Experience Rating and Stochastic Reserving in General Insurance

László Martinek

Doctoral Dissertation

2019

Eötvös Loránd University

Department of Probability Theory and Statistics
Institute of Mathematics

Experience Rating and Stochastic Reserving in General Insurance

Dissertation presented by

László Martinek

to the Eötvös Loránd University
in partial fulfilment of the requirements for the Degree of
Doctor of Philosophy
in the subject of applied mathematics

Doctoral School	Mathematics
Director	Tibor Jordán, Professor, DSc
Doctoral Program	Applied Mathematics
Director	János Karátson, Professor, DSc
Supervisor	N. Miklós Arató, Associate Professor, CSc

Budapest, 2019

Experience Rating and Stochastic Reserving in General Insurance

Doctoral Dissertation, 2019

Supervisor: N. Miklós Arató, Associate Professor, CSc

Eötvös Loránd University

Department of Probability Theory and Statistics

Pázmány Péter sétány 1/C, 1117 Budapest, Hungary

Printed and bound in The Netherlands

Acknowledgement

To start with, I am immensely grateful to Miklós Arató for his dedication as a supervisor. He invested a large amount of effort and time in our collaboration, and he was available for consultation any time, either in person or over the phone. It did not just happen once that we had a long-distance call at 10 pm.

I thank my professors Villő Csiszár, László Márkus, György Michaletzky, Tamás Móri, Vilmos Prokaj, András Zempléni in the Department of Probability Theory and Statistics for the lectures and discussions throughout the years, and for organising seminars and conferences where I could give presentations. Furthermore, thanks to Ágnes Backhausz for the department seminars and for Zsuzsa Vihari and Klaudia Szalay in the Institute of Mathematics for assisting me at any time with all kinds of requests.

It was an unforgettable experience to share a PhD room in the university with Balázs Gerencsér, Tamás Fegyverneki and László Varga, who helped me with fruitful and bright conversations. Special thanks to Miklós Mályusz for his collaboration in one of our papers.

I highly appreciate the funding I received as part of the Apáczai Csere János PhD scholarship TÁMOP 4.2.4. A/1-11-1-2012-0001.

I am extremely grateful to NN Group, where I have been working since 2014, especially Dávid Bozsó and Wieger Postma for encouraging me and for giving me study time and financial support to be able to accomplish this task. I thank Youbaraj Paudel, who showed me that one is capable to go a long way to achieve their goals.

Many thanks to Emma van der Weerd for reviewing the grammar in Chapter 3 from the USA, whilst I was then on the other side of this globe.

Special thanks to Dániel Szakács for designing the cover of the dissertation and for paying attention to my various requests.

I thank so much my mother Edit for her love, encouragement, support and patience over the years. She has dedicated so much energy and time to me to ensure I can study and do research at the university. I also thank my father Péter for his support, as well as my brother Péter for constantly inquiring about the progress of my PhD. My grandparents, Edit and Sándor also deserve my gratitude, and especially Kati mama, who would arguably be the happiest to see this dissertation.

Finally, I am tremendously grateful to Évi for all that she has done to help me complete papers and this dissertation. Her support is not limited to loving patience, but also meant active review of all my draft articles. I have written this dissertation in Amsterdam, Budapest and Tokyo — she has been present in all these locations and has facilitated the progress, not only by uplifting words, linguistic and numeric advice, but also by complaining relatively little about me working until 2 in the morning. I dedicate this work to her.

Contents

1 Introduction

The present dissertation guides the reader through years of research in the field of mathematical applications in non-life insurance. It is separated into two parts: Part I (Chapters 2 and 3) addresses experience rating in the context of bonus–malus systems, whilst the topic of Part II (Chapters 4 and 5) is stochastic claims reserving.

A posteriori ratemaking is widely applied in the premium calculation of general insurance products such as third-party automobile insurance or other property-related products. Premium adjustment is usually done by applying bonus–malus systems, which means that policyholders with a bad claim history pay a surcharge, and policyholders with good history get a discount [59, 61]. The rating system is identical to a random walk on a graph where each vertex is labelled with a premium multiplier, whereby one of the vertices is the predetermined initial one. It is of high importance for the insurer to estimate expected future claims for each insured. Chapter 2 introduces three estimation techniques. We show how slow the convergence to the steady state is in the random walk, and how this time-dependence affects the reliability of claim expectations. Furthermore, we show how the quality of information which is at the insurer's disposal affects the estimations. We connect the competing estimation techniques with the theory of probability scores [46, 45], and apply them for ranking among peer models in an algorithmic way.

Novelties of chapter *2. (1) Estimation of the expected λ number of accidents (claim frequency) triggered by insured drivers by comparing different methods. (2) Parameter estimations (α, β) and their appropriateness based on claim numbers of sample size N. (3) New Bayesian method for the estimtion of λ.*

Chapter 3 proposes an alternative to common frameworks that are designed as random walks on graphs of mostly finite state space which represents premium levels. Several papers have been published that deal with optimal relativities or optimal transition rules, however, these predominantly remain in the realm of bonus–malus systems. The proposed premium calculation model is governed by the policyholder's claim history through a recursive equation. This new autoregressive scheme is structurally different from the ones in use. Relevant metrics that measure the system's optimality are evaluated, partially in analytical forms. Through a comparison with existing models and parameterisation from real-life data, the new model is put into context and its practical relevance is investigated. As a further generalisation compared to the assumptions in the previous chapter, here we assume that the expected claim volume is a stationary Markov chain.

Novelties of chapter *3. (1) Construction of a recursive premium process (experience rating model) which has not been observed in bonus–malus systems before, proof of existence of the stationary distribution. The research is motivated by the search for optimal systems. (2) Analytical formulas that measure process elasticity, coefficient of variation, relative stationary average premium level, financial equilibrium and quadratic loss when the frequency process is a stationary Markov chain. (3) Comparison with real European schemes.*

Chapter 4 turns to stochastic claims reserving [35, 111]. In the past two decades increasing computational power resulted in the development of more advanced claims reserving techniques that allowed the stochastic branch to overcome deterministic methods. This shift resulted in forecasts of enhanced quality and the better understanding of risks borne by the undertaking. Not only point estimates, but predictive distributions can be generated in order to forecast future claim amounts.

The appropriate estimation of outstanding claims is traditionally one of the most important tasks of non-life actuaries. These are claims which have already occurred. The insurance company might learn about some claims with a delay, which can be as considerable as 40 years in extreme cases. However, the company will still have the liability to indemnify the policyholders. For that reason, the amount and appropriateness of the reserves have a substantial effect on the financial results of the institutions. In recent years, stochastic reserving methods have become increasingly widespread, supported by broad actuarial literature, describing development models and evaluation techniques.

This chapter compares the appropriateness of several stochastic estimation methods. For lack of analytical formulas in most of the model settings, simulations are performed to approximate distributions and results. The number of available run-off triangles is fairly limited (insurers do not or do rarely publish them), thus in turn we have defined a simulation-based algorithm for ranking the different reserving techniques. Stochastic reserving is taken into account as a stochastic forecast, thus comparison techniques developed for stochastic forecasts can be applied. Probability integral transform, continuous ranked probability score [47] and further metrics are the key measures of ranking. We elaborate on the interpretation of the exact results.

NOVELTIES OF CHAPTER *4. (1) A simulation-based ranking algorithm of different stochastic reserving models. (2) Apply measures such as CRPS, PIT, coverage, sharpness which are new in claims reserving. Answer questions (a) if these measures are better than regular ones (msep) (b) which methods predict the distributions properly (c) how reliable and sharp the prediction intervals are. (3) Present the ranking framework also on a real data set not used before elsewhere.*

Ultimately, Chapter 5 continues the work from the previous one by extending the research by establishing new models and by applying the comparisons on real-life data from more than 700 institutions. The significant expansion in the variety of models requires that we validate these methods and define appropriate decision making. This chapter compares and validates several existing and self-developed stochastic methods on actual data, applying comparison measures in an algorithmic manner. The concept 'experience rating' in this chapter relates to the credibility methods in contrast to the no-claim discount in Part I.

NOVELTIES OF CHAPTER *5. (1) We present new metrics in actuarial reserving such as CRPS, coverage and sharpness of several models, to analyse their performance and determine an order of appropriateness. (2) PIT has already been applied by [77] on stochastic models, here we continue presenting the calculations involving further methods not covered elsewhere (credibility bootstrap, bootstrap Munich, semi-stochastic). (3) Two new models are introduced: credibility bootstrap and collective semi-stochastic. (4) We emphasise the importance of an algorithmic way of model selection from competing peers. Model performance assessment has been made on* actual data*. (5) Models based on internal information only (single triangle) are also compared with collective ones (multiple triangles and credibility), and the present chapter intends to convey the potential of oversight data collection and possible application on multiple triangles. (6) Scripts have been made available online.*

For each of the chapters a considerable amount of self-developed scripts have been programmed in `R`. All calculations and algorithms in this dissertation have been implemented in this language. In addition, packages `ChainLadder`, `ggplot2`, `rjags`, `xtable` have been exploited as well. These program codes have supported the drafting of conjectures and the creation of tables and figures.

The papers of the author embedded into the dissertation are

- [3, 73] in Chapter 2,
- [73] in Chapter 3,
- [2, 4, 72] in Chapter 4,
- [72] in Chapter 5.

Part I

Experience Rating in General Insurance

You can know the name of that bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird. You'll only know about humans in different places, and what they call the bird. So let's look at the bird and see what it's doing—that's what counts.

— **Richard P. Feynman**

Schemes and Claim Number Estimation in Automobile Insurance

2

2.1 Experience rating

Premium calculation for several types of general insurance[1] products requires the consideration of former observations. Insurance schemes require policyholders to pay based on their level of risk, rather than based on forming a unified risk community. The incorporation of each policyholder's underlying risk into the price allocation scheme can simultaneously be based on the segmentation of population behaviour and on the individual experience. In automobile liability insurance, *a priori* ratemaking connects to the former and applies generalised linear models to certain policy characteristics, such as motor power, CO_2 emission, driver age and gender, territory and purpose of vehicle use. A wide range of literature is available discussing these factors extensively, see [59]. The premium calculation is often realised involving credibility techniques, mixing the insurance institution's experience with available national data, applying an appropriate proportion. For the discussion of credibility see [41, 28].

Let claim frequency stand for the theoretical expectation of the number of claims filed by a policyholder. The supposed a priori rating suggests a frequency parameter, which is later adjusted by the individual claim history as an a posteriori multiplicative premium adjustment coefficient, thereby correcting the imprecision of the a priori rating. A possible representation of this a posteriori rating is the bonus–malus system, see [61] for a brief introduction. This type of merit rating, also referred to as an experience rating system or occasionally as a no-claim discount system, should not necessarily be constrained to motor third-party liability insurance. Rather, this rating may also be applied in other types of property and casualty insurance where the observation of individual claim history is a relevant part of the premium calculation. The term *bonus* signifies a discount in the price of the insurance coverage if the

[1] Identical terminologies to general insurance are *property and casualty insurance*, mainly used in the United States and Canada, and *non-life insurance*, used in Continental Europe.

policyholder has a clean record, whereas the term *malus* refers to the surcharge in premium costs for records containing loss reports. A generalised bonus–malus system combining both a priori and a posteriori information can be studied in [30, 106].

In compulsory liability automobile insurance the bonus–malus system is applied worldwide, especially in Europe and Asia. All member states of the European Union apply no-claim discount systems of diverse characteristics. These schemes are constructed in order to support the segmentation of drivers in premium calculation, by penalising the more dangerous drivers with higher contributions. It means that policyholders with bad history should pay more than others without accidents recorded. Throughout this chapter we will apply this terminology from the perspective of automobile insurance, however, the concept bonus–malus has a wider spectrum of applications. Experience rating has been implemented among others in medical malpractice insurance, unemployment insurance and workers' compensation insurance in several countries. The benefits in each case are not underpinned by statistical evidence. The effect of such a system in medical malpractice to reduce the negligence of third parties through the threat of a higher premium has been challenged by [99].

Existing bonus–malus schemes are identical to random walks on a finite state space of premium levels, often called relativities. Transition rules regulate the subsequent step from one state to another after a period with k claims on the policyholder's account. As a general rule respected by most existing systems, premiums are reduced or do not increase after a claim-free period, and they do not decrease when a claim is reported. A thorough elaboration of bonus–malus schemes can be found in [60].

The underlying system can be interpreted as a graph with vertices called classes, responsible for the a posteriori premium adjustment as a multiplier. Each new driver begins in the initial vertex by definition, motivated by the fact that apart from the a priori features, every insured is handled equally due to the lack of a historical record with respect to claims. In the majority of practical implementations, the graph contains a finite number of classes, however, infinite systems do exist as well, see [59]. Having spent a year without causing any accident, the insured person's relativity level migrates into another class, associated with a lower premium. Otherwise, in case of an accident, the insured shifts downward on the bonus scale to a new class with higher premium, unless he or she was already in the worst one, provided that the number of classes is finite and there is a maximum level indeed. By convention, let the lowest class stand for the lowest premium (highest bonus) and the highest class for the highest premium (highest penalty).

In this chapter the focus is on the bonus–malus systems, i.e. the a posteriori ratemaking and the information it provides on the expected number of claims per policyholder. Our aim is to estimate the expected λ number of accidents caused by insured drivers. This expectation is usually referred to as the claims frequency of the policyholder. In section 2.2 we review the required assumptions, among others the distribution of claim counts of a policyholder in one year or any period, and the Markovian property of a random walk in such a system of relativities.

Section 2.3 explains three alternative manners of calibrating the population parameters, defining an a priori distribution. Section 2.4 elaborates on estimation problems with three models proposed, illustrated with Belgian, Brazilian and Hungarian examples. Each of the methods assume the existence of certain information about the driver, which includes the claims history and the bonus class. By taking a Bayesian approach, the first method creates the a posteriori distribution of claim frequency given the last bonus class. Fitting the best possible estimation for λ is a crucial task for the insurer, since the expected value of claims, and more importantly, the claims reimbursed by the insurance firm, are forecast using λ. To evaluate the size of benefit payments, the size of the property damage has to be approximated, not only the number of them. This part of actuarial calculations is not taken into account in the present chapter, for further discussions see [40, 70]. Section 2.5 describes score definitions used for the ranking of competing models. These metrics are used in section 2.6, by determining an algorithm to compare the three methods. The score-based comparison is more general, as this will be shown in later chapters. Hence, in this chapter a comparison method of different claim frequency estimation models is created, using various information, based on [3] and [73].

2.2 Assumptions

In this section a few assumptions are specified with respect to the claim number distribution, the policyholder's riskiness and the transition from one premium class to another.

Assumption 2.1 (claim number) The policyholder's claim number for a year or a fixed period of time is Poisson(λ).

The Poisson assumption has been challenged by [112], proposing zero-inflated models as alternatives, such as zero-inflated negative binomial, zero-inflated generalised Poisson or zero-inflated Delaporte distributions, due to the dispersion effects observed on actual claim reports. Nevertheless, these distributions cannot be interpreted as standalone alternatives, but are closely linked to the next assumption about the λ

parameter itself. This one reflects the intuition that the λ frequency strongly depends on the insured person's character, which can stem from a finite or infinite set of riskiness categories.

Assumption 2.2 (mixing distribution) Suppose that the claims frequency parameter is also a random variable, denoted by Λ. It is governed by the *mixing distribution*, which is chosen to be the Gamma mixing distribution, as distribution commonly applied in the literature.

In other words, a person randomly drawn from the population has a frequency parameter governed by the mixing distribution. The gamma distributed case is discussed in the present chapter, as most commonly assumed by a wide range of papers. It does not constrain generality. For other assumptions the calculations may become more complicated.

Definition 2.1 (Gamma distribution) Let $\Gamma(\alpha, \beta)$ be the gamma distribution with $\alpha > 0$ shape and $\beta > 0$ rate parameters, and with density function $f(x) = \frac{x^{\alpha-1} \cdot \beta^{\alpha} \cdot e^{-\beta x}}{\Gamma(\alpha)}$ for $x > 0$. (Alternatively, with scale parameter $1/\beta$.)

Definition 2.2 (Negative binomial distribution) Let $NB(r, p)$ stand for the negative binomial distribution, where $r \in \mathbb{Z}_+$ is the number of failures until the last experiment and $p \in (0, 1)$ is the probability of success. If $\xi \sim NB(r, p)$, then $P(\xi = k) = \binom{r+k-1}{k} p^r (1-p)^k$, where $k = 0, 1, \ldots$. (Parameter r can be extended to $r \in \mathbb{R}_+$.)

On the condition that $\{\Lambda = \lambda\}$, the number of property damages related to a driver is a random variable $X \sim \text{Poisson}(\lambda)$, as already specified that the conditional probability is $P(X = k | \Lambda = \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$. Assuming gamma mixing variable, it is known that the unconditional distribution is negative binomial with parameters $\left(\alpha, \frac{\beta}{1+\beta}\right)$, see [28]. Let $\Lambda \sim \Gamma(\alpha, \beta)$. Accordingly, the a priori parameters can be estimated either by a standard method of moments or by the maximum likelihood method. Certainly, it is necessary to perform hypothesis testing, as the assumptions regarding the mixing distribution might be inaccurate.

Besides the gamma mixing distribution, other alternatives of claims frequency distributions might be worth considering. If Λ is Inverse Gaussian, the unconditional distribution of X is Poisson Inverse Gaussian, see [39, 105]. Furthermore, the case of $\Lambda \sim$ Log-normal is also a realistic option, see [39]. For a more general parametric consideration see [108], which contains Poisson, Negative Binomial or Poisson Inverse Gaussian distributions as special cases, however, it requires 3 parameters. Besides the parametric assumptions, the non-parametric estimation is also worth considering, see [27].

Assumption 2.3 (independence) λ is a constant value in time for each policyholder as the realisation of a Λ random variable. Furthermore, suppose that the Λ variables of the entire population are independent.

The case of time-dependent $\Lambda(t)$ implies double stochastic processes, such as Cox processes, i.e. $\Lambda(t)$ would also be a stochastic process itself. For a time-dependent analysis and dependence between contracts see [12].

The following assumption differs from the former ones, as it specifies the randomness of premium relativities instead of the claim counts.

Assumption 2.4 (homogeneous Markov chain) The random walk on the graph of classes is a homogeneous Markov chain, i.e. each subsequent step depends only on the last state and it is homogeneous in time.

Notation 2.1 Consider a bonus–malus system consisting of n premium classes. Let C_1 denote the best relativity with the highest bonus, C_2 the second best one, and finally, let C_n be the worst achievable class with the highest premium penalty. Moreover, let Y_t be the class of the policyholder after t steps (years).

In terms of these notations, the Markovian property can be written as $P(Y_t = C_i|Y_{t-1}, \ldots, Y_1) = P(Y_t = C_i|Y_{t-1})$. As the Y_t process is supposed to be homogeneous, let p_{ij} stand for $P(Y_{t+1} = j|Y_t = i)$. These values specify an $n \times n$ stochastic matrix with non-negative elements, which is the transition probability matrix of the random walk on the set of states. Let us denote it by $M(\lambda) \in \mathbb{R}^{n\times n}$. Now we enumerate three different schemes from Belgium, Brazil and Hungary. The primary reason for selecting these models is that the Brazilian one consists of a low number of relativities, whilst the Belgian one is sophisticated in the sense that it contains a larger amount of classes. The Hungarian system is not extreme from that perspective, having an average number of relativities, see Table 2–1[2]. However, the international literature lacks papers with focus on the Hungarian scheme, hence the choice.

Example 2.1 (Hungary) In the Hungarian bonus–malus system there are 15 premium classes, an initial one (A_0), 4 malus ($M_4, \ldots, M_1$) and 10 bonus ($B_1, \ldots, B_{10}$) classes. Using the aforementioned notations, we can think of it as $C_{15} = M_4, \ldots, C_{12} = M_1, C_{11} = A_0, C_{10} = B_1, \ldots, C_1 = B_{10}$. After every claim-free year the policyholder

[2]Source: Belgium [62], Brazil [62], China [81], Hungary [24], Japan [75] (modified according to [81]), Korea [81], Netherlands [37].

	premium relativities $(C_n, \dots, C_1)$	number of classes
Belgium	2, 1.6, 1.4, 1.3, 1.23, 1.17, 1.11, 1.05, 1, 0.95, 0.9, 0.85, 0.81, 0.77, 0.73, 0.69, 0.66, 0.63, 0.6, 0.57, 0.54, 0.54, 0.54	23
Brazil	1, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65	7
China	1.3, 1.2, 1.1, 1, 0.9, 0.8, 0.7	7
Hungary	2, 1.6, 1.35, 1.15, 1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5	15
Japan	1.5, 1.4, 1.3, 1.2, 1.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.45, 0.42, 0.4, 0.4, 0.4	16
Korea	2, 1.9, 1.8, 1.7, 1.6, 1.5, 1.4, 1.3, 1.2, 1.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.45, 0.4	18
Netherlands	1.15, 0.95, 0.85, 0.75, 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.28, 0.25	14

Table 2–1: Relativities in a number of bonus–malus systems.

moves one step down, unless he or she was in B_{10}, when there is no better class to go to. The consequence of every reported damage claim is an upgrading of 2 classes, and at least 4 damage claims pull the driver back to the worst M_4 state. Thus the transition probability matrix takes the form of Equation B.13, see Appendix B.

Example 2.2 (Brazil) 7 premium classes: $A_0(\text{initial class}), B_1, B_2, \dots, B_6$. Sometimes written as classes $7, 6, 5, \dots, 1$. Transition rules can be found in [61].

Example 2.3 (Belgium) The current (new) Belgian system was introduced in 1992. We address the transition rules regarding business-users[3], which can be found in [61]. There are 23 premium classes: M_8, M_7, $\dots$, M_1, A_0, B_1, B_2, $\dots, B_{14}$ (sometimes indicated as classes $22, 21, \dots, 1, 0$).

2.3 Estimation of mixing distribution parameters

In this section we construct estimation methods to compute the approximate values of parameters α and β. The first two methods are modifications of the method of moments and the last one follows the maximum likelihood method. Suppose that the insurance company maintains claim statistics from the past few years containing m policyholders. The ith insured person caused X_i accidents by his or her fault[4]

[3] Non-business-users enter the system in C_{11}, business users enter in C_{14}. By convention, the highest bonus class is denoted by C_0.

[4] In an accident that involves multiple parties, the one who was responsible for the outcome bears the liability to indemnify the others. (Exclude highly complicated cases.) This is the liability which is forwarded to the insurer and which affects on his or her claim history.

over a time period of t_i years, where t_i is a positive real number. Indeed, t_i is not required to be an integer, because any person may switch insurer in the middle of a policy period or other circumstances may arise, as a result of which the observation may correspond to a fraction of a year. According to the present assumptions, X_i is governed by Poisson$(t_i \cdot \Theta)$, where exogenous variable t_i is a personal time factor and Θ is a Gamma(α, β)-distributed random variable. The unconditional distribution of X_i as mentioned above is Negative Binomial$\left(\alpha, \frac{\beta}{t_i+\beta}\right)$, thus $E(X_i) = \frac{\frac{t_i}{t_i+\beta}\alpha}{1-\frac{t_i}{t_i+\beta}}$ and $Var(X_i) = \frac{\frac{t_i}{t_i+\beta}\alpha}{\left(1-\frac{t_i}{t_i+\beta}\right)^2}$ imply that the first two moments are

$$EX_i = t_i\frac{\alpha}{\beta} \quad \text{and} \quad EX_i^2 = t_i^2\frac{\alpha}{\beta^2} + t_i^2\frac{\alpha^2}{\beta^2} + t_i\frac{\alpha}{\beta}. \tag{2.1}$$

Since the method of moments implies several corresponding systems of equations, there is a freedom in the choice of selecting one of them. On the basis of the moments above, $\sum\limits_{i=1}^{m} EX_i = \frac{\alpha}{\beta}\sum\limits_{i=1}^{m} t_i$ and $\sum\limits_{i=1}^{m} EX_i^2 = \left(\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}\right)\sum\limits_{i=1}^{m} t_i^2 + \frac{\alpha}{\beta}\sum\limits_{i=1}^{m} t_i$, thus the estimators of parameters are the result of the following equations.

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{\sum\limits_{i=1}^{m} X_i}{\sum\limits_{i=1}^{m} t_i} \tag{2.2}$$

$$\frac{1}{\hat{\beta}} = \frac{\sum\limits_{i=1}^{m} t_i}{\sum\limits_{i=1}^{m} t_i^2}\left(\frac{\sum\limits_{i=1}^{m} X_i^2}{\sum\limits_{i=1}^{m} X_i} - 1\right) - \frac{\sum\limits_{i=1}^{m} X_i}{\sum\limits_{i=1}^{m} t_i} \tag{2.3}$$

Definition 2.3 (MM1) Let the first method of moments estimation be defined by Equation 2.2 and Equation 2.3.

A slight modification in the formulas result in $\sum\limits_{i=1}^{m} \frac{EX_i}{t_i} = m\frac{\alpha}{\beta}$, thus the first equation for the estimators changes to

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{\sum\limits_{i=1}^{m} \frac{X_i}{t_i}}{m}. \tag{2.4}$$

Definition 2.4 (MM2) Let the second method of moments be defined by the solution of Equation 2.4 and Equation 2.3.

Generally, MM1 and MM2 do not provide the same results, except for $t_1 = t_2 = \ldots = t_m$. Numerous claims history scenarios were simulated for several portfolios, and based on our experience, the solutions of MM1 provided the better estimations for α and β. As a third method, the maximum likelihood estimator is considered. However, the likelihood function has no maximum in a significant amount of cases of practical relevance.

Let $\underline{t} = \begin{pmatrix} t_1 & \dots & t_m \end{pmatrix}^\top$ and $\underline{X} = \begin{pmatrix} X_1 & \dots & X_m \end{pmatrix}^\top$ denote the period length and claim number observations. The likelihood functions is

$$L(\alpha, \beta, \underline{t}, \underline{X}) = \prod_{i=1}^{m} \frac{\Gamma(X_i + \alpha)}{\Gamma(\alpha) \cdot X_i!} \cdot \left(\frac{\beta}{t_i + \beta}\right)^{\alpha} \cdot \left(\frac{t_i}{t_i + \beta}\right)^{X_i}, \tag{2.5}$$

thus the loglikelihood is

$$\begin{aligned} \log L(\alpha, \beta, \underline{t}, \underline{X}) = & \sum_{i=1}^{m} \log \Gamma(X_i + \alpha) - \log \Gamma(\alpha) - \log X_i! + \alpha \log \beta - \\ & - \alpha \log(t_i + \beta) + X_i \log t_i - X_i \log(t_i + \beta). \end{aligned} \tag{2.6}$$

The solution of the system of equations $\frac{\partial \log L}{\partial \alpha} = 0$ and $\frac{\partial \log L}{\partial \beta} = 0$ provides the $\hat{\alpha}_{ML}, \hat{\beta}_{ML}$ maximum likelihood estimators, where

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{m} \psi^0(\alpha + X_i) - m\psi^0(\alpha) + m \log \beta - \sum_{i=1}^{m} \log(t_i + \beta), \tag{2.7}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{m\alpha}{\beta} - \alpha \sum_{i=1}^{m} \frac{1}{t_i + \beta} - \sum_{i=1}^{m} \frac{X_i}{t_i + \beta} = \alpha \left(\frac{m}{\beta} - \sum_{i=1}^{m} \frac{1}{t_i + \beta}\right) - \sum_{i=1}^{m} \frac{X_i}{t_i + \beta} \tag{2.8}$$

and $\psi^0(x) = \frac{\frac{\partial}{\partial x}\Gamma(x)}{\Gamma(x)}$ denotes the digamma function.

Definition 2.5 (ML) Let the maximum likelihood estimator $\hat{\alpha}_{ML}, \hat{\beta}_{ML}$ be defined by the solution of Equation 2.7 and Equation 2.8 for 0.

	mse MM1	mse MM2	mse ML	MM1 fails	MM2 fails	ML fails	MM1 best	MM2 best	ML best
α	51.5	53.1	1.2	30%	30%	30%	20%	20%	60%
β	695.3	861.8	19.3	30%	30%	30%	20%	20%	60%

Table 2–2: Estimated alpha and beta parameters from a sample of size 100. The number of simulations is N = 5000.

Approximating parameters from a sparse sample or from a larger one affects the performance of the three methods, see Table 2–2, Table 2–3, Table 2–4, Table 2–5, analyses on a spectrum of 100 to 5000 policyholders. Original parameters are

	mse MM1	mse MM2	mse ML	MM1 fails	MM2 fails	ML fails	MM1 best	MM2 best	ML best
α	408.4	86.5	1.1	0%	0%	10%	20%	30%	50%
β	6921.9	1414	17.5	0%	0%	10%	30%	30%	50%

Table 2–3: Estimated alpha and beta parameters from a sample of size 500. The number of simulations is N = 5000.

	mse MM1	mse MM2	mse ML	MM1 fails	MM2 fails	ML fails	MM1 best	MM2 best	ML best
α	23.1	6.1	0.8	0%	0%	0%	30%	30%	40%
β	365.7	98.4	13.2	0%	0%	0%	30%	30%	40%

Table 2–4: Estimated alpha and beta parameters from a sample of size 1000. The number of simulations is N = 5000.

$\alpha = 1.2, \beta = 19$. In each table, columns 1-3 stand for the squareroots of mean square error (mse) values.

Definition 2.6 (MSE 1) Let the *mean square error* associated with parameter ϑ be $mse(\vartheta, \text{method}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\vartheta}_{method,i} - \vartheta)^2$, where ϑ is the real value and $\hat{\vartheta}_{method,i}$ is the approximation from the i^{th} sample (not the size of one sample). Parameter ϑ stands for either α or β in the calculations presented, and 'method' in the subscript stands for either MM1, MM2 or ML.

Columns 4-6 include the proportion of cases when a method fails to make any prediction, such as showing negative results or the lack of solution in the case of the maximum likelihood equations. Columns 7-9 show the proportions when the methods result in the most accurate estimations, i.e. being the closest to the real parameter in absolute value.

One further assumption should be made in the simulations with respect to the exogenous t_i values, as their presence makes the comparison different from regular negative binomial estimations. Assume that each t_i is governed by $Unif(1, 5)$, i.e. a randomly selected policyholder's observation falls between 1 and 5 years. Recall that equal t_is make MM1 and MM2 identical.

MM1 fails with the lowest chance and for a larger sample size the result is comparable to the one given by the maximum likelihood method. We can conclude that for a population of at least 5000 policyholders MM1 provides a reasonable solution.

We give a second definition for the mean square error, which captures the approximation error through the difference of probabilities based on real ϑ and estimated $\hat{\vartheta}$ parameters.

	mse MM1	mse MM2	mse ML	MM1 fails	MM2 fails	ML fails	MM1 best	MM2 best	ML best
α	0.3	0.3	0.3	0%	0%	0%	30%	30%	40%
β	4.6	4.8	4.2	0%	0%	0%	30%	30%	40%

Table 2–5: Estimated alpha and beta parameters from a sample of size 5000. The number of simulations is N = 5000.

	sample size: 100	sample size: 500	sample size: 1000	sample size: 5000
MM1	0.0352	0.0165	0.0118	0.0053
MM2	0.0385	0.0184	0.0133	0.0060
ML	0.0351	0.0163	0.0116	0.0053

Table 2–6: Estimated squareroot of MSE of the methods from different population sample sizes, according to Definition 2.7. The number of simulations in each case is N = 5000.

Definition 2.7 (MSE 2) Given the sample $x_1, x_2, \ldots, x_m$, let mse be

$$E\left(\frac{1}{m}\sum_{i=1}^{m}\left(P_{\hat{\vartheta}_i}(x_i) - P_{\vartheta_i}(x_i)\right)^2\right),$$

where $\vartheta_i = \left(\alpha, \frac{\beta}{t_i+\beta}\right)$ and $\hat{\vartheta}_i = \left(\hat{\alpha}_{method}, \frac{\hat{\beta}_{method}}{t_i+\hat{\beta}_{method}}\right)$. Table 2–6 shows the performance of estimation models in line with Definition 2.7. The expectation is approximated by the average of 5000 independent simulations of $\frac{1}{m}\sum_{i=1}^{m}\left(P_{\hat{\vartheta}_i}(x_i) - P_{\vartheta_i}(x_i)\right)^2$. MM1 and ML are nearly identical and they outperform MM2, whilst MM1 provides a solution with a higher chance, as discussed above.

2.4 Methods of claim frequency evaluation

In this section we introduce three methods to evaluate policyholders' claim frequencies. Each of these models concentrate on the predictive power of bonus–malus classes one way or another. In other words, from the perspective of claim frequency evaluation, the question is how powerful the information can be regarding the last class of the insured, and to what extent accurateness changes if the complete history is known. It is crucial to assess the implications of the lack of specific data. Another question is how time affects the predictions, in the sense that the policyholder has spent 5, 10, 20 or even more years in the system. Convergence to stationary states can be far beyond the general insurance coverage periods. The three methods discussed in the present section are as follows.

method 1 Time t spent in the scheme and last class, let this be denoted by $Y_t = c$. The value of time is given in years and c is one of the premium classes

$C_1, \ldots, C_n$.

method 2 The policyholder's current class. Claim statistics of the insurance institution from a previous year.

method 3 Time t spent in the scheme and total class history.

Furthermore, each of the models use portfolio data for the evaluation of the mixing distribution parameters, see Section 2.3. Table 2–7 summarises the model conditions. All the other assumptions made previously in this chapter remain unchanged.

	years in the system	premium class	external data
method 1	known	the last one	claim observations of a portfolio, used for α, β
method 2	not known	the last one	claim observations of a portfolio, used for α, β and for the average claims per class
method 3	known	total history	claim observations of a portfolio, used for α, β

Table 2–7: Summary of policyholder information used in the claim frequency estimation models.

2.4.1 Conditional probabilities and the latest class

Insureds often migrate from one insurance institution to another, which is usually triggered by more favourable prices. When an insured person changes insurance institution, the new company may not necessarily get his or her entire claim history, only the class where his or her life has to be continued. In practice, this is primarily due to poor data governance or lack of integrated systems. The new insurer also knows the number of years the policyholder has spent in the liability insurance system. Nevertheless, based on the information *years* and *last class*, the current model is designed to provide the best possible estimation for the person's λ.

Recall that $\{Y_t = c\}$ denotes the event that the investigated policyholder has spent t years in the system and arrived in class c. More precisely, from the initial class the number of movements (steps) is t.

Definition 2.8 (transition probability matrix) Let matrix $M(\lambda) \in M_n(\mathbb{R})$ denote the transition probability matrix of the random walk on the states of premium classes. See Equation B.13.

Definition 2.9 (initial distribution) π_0 is the initial discrete distribution, which is a column vector of the form $(0, \ldots, 0, 1, 0, \ldots, 0)^\top$. The ith element is 1, which means that each driver begins in the initial C_i state.

Assume that the a priori α, β parameters are evaluated. Consider the information $\{Y_t = c\}$ besides the initial class Y_0, which is fixed for each insured person. According to Bayes' theorem, the conditional density of Λ is

$$f_{\Lambda|Y_t=c}(\lambda|c) = \frac{P(Y_t = c|\Lambda = \lambda) \cdot f_\Lambda(\lambda)}{\int\limits_0^\infty P(Y_t = c|\Lambda = \lambda) \cdot f_\Lambda(\lambda) d\lambda}, \tag{2.9}$$

and the estimation for λ is the a posteriori expected value $\hat{\lambda} = \int\limits_0^\infty \lambda \cdot f_{\Lambda|Y_t=c}(\lambda|c) d\lambda$. There is no closed formula for $P(Y_t = c|\Lambda = \lambda)$, which can only be evaluated pointwise as a function of λ, calculating the tth power of the transition matrix $M(\lambda)$. If I denotes the index of the initial class in the graph, this probability is exactly the $M(\lambda)^t_{(I,|c|)}$ element of the matrix, where $|c|$ is the index of class c. Using numerical integration, $\hat{\lambda}$ can be computed relatively fast. Let this method be referred as *method 1*. Parameters $\alpha = 0.8888$ and $\beta = 6.0299$ in the calculations have been taken from a real Belgian automobile portfolio, see [86]. Note that the original paper uses a different parameterisation such that the claim frequency is $\lambda\vartheta$ with a fixed $\lambda = 0.1474$ and $\vartheta \sim \Gamma(a, a)$, $a = 0.8888$. Figure 2–1, Figure 2–2 and Figure 2–3 present the claim frequency estimations in the countries observed. For the sake of transparency, only 6 classes are represented on each plot. Each point is the approximation of $\hat{\lambda}$, given the current class and years spent in the scheme. Connected lines represent classes and provide evidence how the a posteriori value of λ can change over time. For instance, a Brazilian driver in class C_5 after 2 years is estimated to have an average claim count of approximately 0.1, whilst this value after 30 years becomes 0.6, see Figure 2–2. This behaviour of the system is determined by the speed of convergence to a stationary distribution on the classes. It takes considerable amount of time for any α, β values of practical relevance to approach the stationary state. Many decades may elapse until the stability is reached. Implicitly, this laziness indicates that any measure of system which focusses on stationarity needs to be taken with caution. Thus, the elapsed time of the individual in the system is important. The meaning of a partially (dis-)connected plot in a class is that the probability of being in that class after t years is zero, see pattern $C8$ in figure 2–1.

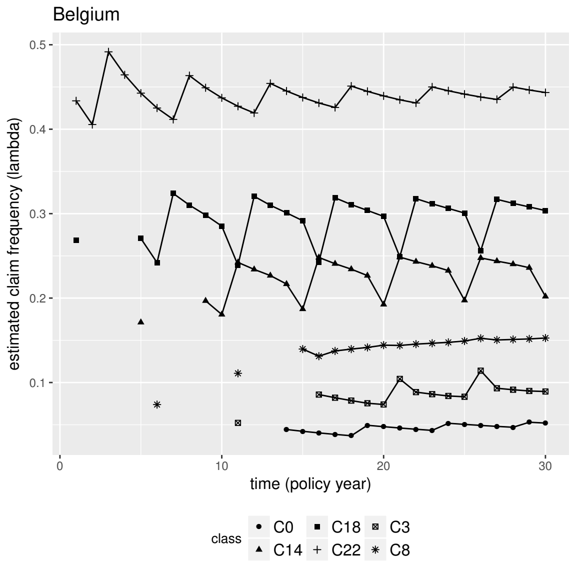


Figure 2–1: Estimated λ parameters as a function of elapsed years and current class of the individual in Belgium (method 1).

2.4.2 Average claim numbers of classes

Take the claim observations concerning a previous period (year) in each class from the portfolio of policies available. This can be interpreted as an in force set of policies, also used for the mixing distribution's parameter estimation. Let *population* stand for this portfolio, *previous year* to a former coverage period in the sense that it precedes the current period with one step. Any former year can be taken, provided that the claim distributions in each of the classes are time-invariant. Certainly, this invariant property requires that the portfolio is homogeneous, i.e. no substantial change can be observed in the behaviour of drivers, the system converged sufficiently in the past. The estimator of a policyholder's claims frequency, who is in class C_i is defined as the average number of claims related to the population in bonus class C_i in the previous year, i.e.

$$\hat{\lambda}_{C_i} = \frac{\sum\limits_{j=1}^{m} k_j \cdot \chi\{j\text{th policyholder is in class } C_i\}}{\sum\limits_{j=1}^{m} \cdot \chi\{j\text{th policyholder is in class } C_i\}}, \tag{2.10}$$

where m is the total number of policies in previous year, k_j is the claim number regarding the jth person, and χ is an indicator function. We refer to this model as *method 2*.

Suppose that the institution's portfolio contained 5 policyholders in the previous year, each in class C_i with claims $0, 1, 0, 0, 2$. Then the estimation for the insureds'

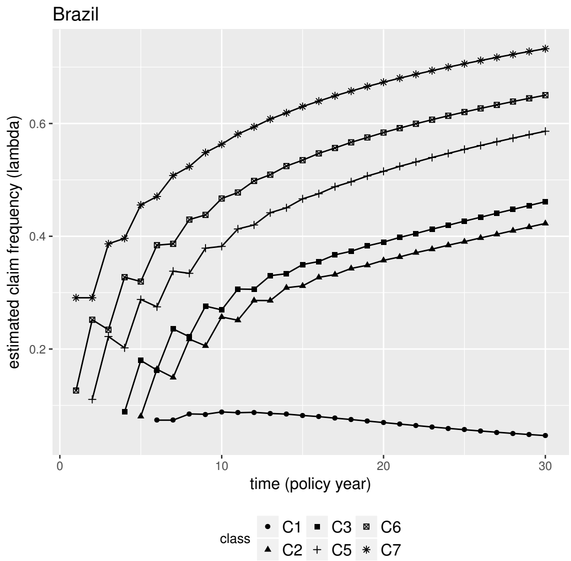


Figure 2–2: Estimated λ parameters as a function of elapsed years and current class of the individual in Brazil (method 1).

frequencies in the current year in class C_i is $\hat{\lambda}_C = 0.6$. Despite the fact that this is a fairly simple model, under specific circumstances it outperforms the other two models. The explanatory power of *method 2* is discussed later in this chapter.

2.4.3 Claim history of individuals

Suppose that the entire claim history of the individuals is available. This can be the case either if the policyholder has previously been insured by the current insurance institution, or if the claims data have been transferred successfully from one company to another. The assumption that yearly claim numbers are independent and identically distributed still holds. Let the insured's history be of length $t \in \mathbb{R}$ years and let X denote the number of aggregate claims caused by this person in the t-year-long presence in the system. Then the conditional distribution of X is also Poisson with parameter λt. Firstly, the α and the β parameters have to be estimated exactly the way we have seen it before in section 2.3. The estimation of λ is the conditional expected value of the Gamma distributed Λ given $\{X = x\}$, i.e. $\hat{\lambda} = E(\Lambda | X = x) = \frac{x+\alpha}{t+\beta}$.

This third method is generally prevalent in the actuarial practice. Refer to it as *method 3*. This is also a Bayesian approach similarly to *method 1*, but the condition is different, which assumes a broader range of knowledge.

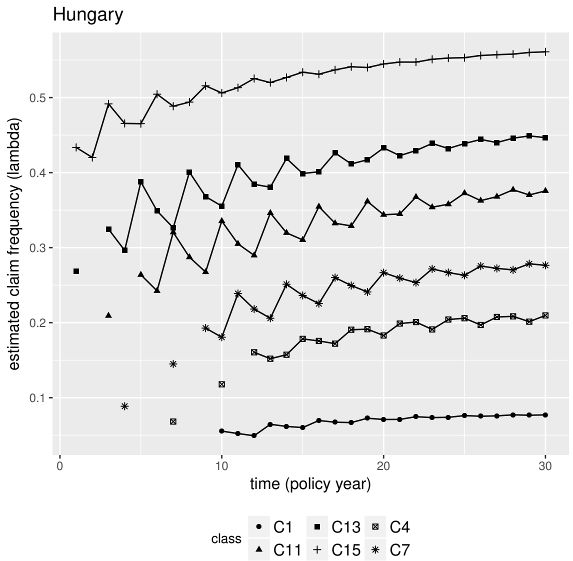


Figure 2–3: Estimated λ parameters as a function of elapsed years and current class of the individual in Hungary (method 1).

2.5 Comparison of estimation methods with scores

The aim of this section is to introduce score metrics assessing the performance of estimation methods described in section 2.4. However, it is equally important to define the concept of ranking in general as the actual detailed performance produced by the three methods on the Belgian, Brazilian and Hungarian examples. Score ranking can be used for the comparison of any other methods and underlying parameters. Scores provide a powerful tool for probabilistic forecasting implemented in multiple disciplines, such as finance, meterorology, insurance or medicine.

A comprehensive overview of the characterisation of proper scoring rules, scores for categorical, continuous variables and beyond is provided in the highly referred [46]. Evaluation from a point forecast perspective can be found in [45], as in many circumstances of practical relevance, the decision has to be made about a single-point-valued forecaster. The paper illustrates the competing entities with a statistician, an optimist and a pessimist forecaster, where the optimist counterintuitively scores the best using the absolute error. This finding highlights the fact that further metrics have to be seen for a reasonable comparison instead of the common scoring rules such as absolute or squared error.

In [91] scoring ignorance is applied to measure forecast quality, illustrated with a meteorological case study concerning temperature data. In relation to mortality rates, model prediction from a medical aspect is analysed in [55], using the logarithmic score and the deviance information criterion (DIC). Several types of scores, the continuous ranked probability score and its threshold-weighted modification, the logarithmic score, the Brier score, the conditional likelihood score, the censored likelihood score are demonstrated in [65] along with an econometric case study with US GDP growth data. [56] is also an example of financial data series aspect, determining a combination of weights for density forecasts.

In the subsequent chapters of the dissertation the relevant theory is further elaborated; the present section is limited to the pursuit of claim frequency estimation accurateness.

Scores are designed to measure the accuracy of probabilistic forecasts, i.e. to measure the goodness-of-fit of evaluations concerning future events or hidden variables. Technically, they define a ranking among competing predictors, similarly to utilities, where the maximum expected utility is the most beneficial one. The predictive distribution can be represented by its distribution function, empirical distribution function or probability density function, depending on the type of score applied in the specific case. Forecasts are then ranked by comparing the average score of the predictive distributions from each model.

Definition 2.10 (scoring rule) Let Ω be a sample space, $\mathcal{A}$ a σ-algebra of subsets of Ω and let $\mathcal{P}$ be a family of probability measures on $(\Omega, \mathcal{A})$. The scoring rule is the $S(F, x) : \mathcal{P} \times \Omega \to \overline{\mathbb{R}} = [-\infty, \infty]$ real valued functional with the two possible exceptions of $-\infty$ and $+\infty$. The first argument can be interpreted as a prediction, whilst the second one as a realisation.

If P stands for a predictive probability measure and x for an event that materialised, then $S(P, x)$ is intuitively the reward for the prediction made, provided that x turned out to be the real observation. Let there be an estimation denoted by P and the real measure Q governing the phenomenon that is scrutinised. Having defined a scoring rule, the value to find is the average of these scores according to measure Q.

Definition 2.11 (expected score) Let the expected score be

$$S(P, Q) = \int S(P, \omega) dQ(\omega),$$

where measure P is an estimation and Q stands for the real one. Alternatively, $S(P, Q) = E_Q(S(P, X))$.

With $S_n(\vartheta) := \frac{1}{n} \sum_{i=1}^{n} S(P_\vartheta, X_i)$ the estimator $\hat{\vartheta}_n = \arg\max_{\vartheta} S_n(\vartheta)$ is called the optimum score estimator. To put it into perspective, the following relation can be established:

$$\text{maximum likelihood estimators} \subset \text{optimum score estimators} \subset \text{M-estimators}.$$

Without the loss of generality, suppose that forecast P_1 is not worse than P_2, if $S(P_1, x) \geq S(P_2, x)$ in expectation, where x is governed by probability measure Q. Two important features of scores are as follows. The first one ensures that the applied mapping defined as score results in an extreme value if the prediction coincides with the predicted distribution, i.e. the best reward results from the perfect forecast. The second one makes the rules used real valued, implying well-ordering on the range of possible scores related to predictions.

Definition 2.12 (proper scoring rule) Scoring rule S is *proper* relative to $\mathcal{P}$ if $S(Q,Q) \geq S(P,Q)$ for all $P, Q \in \mathcal{P}$. S is *strictly proper* if $S(Q,Q) = S(P,Q)$ if and only if $P = Q$. See [10, 53].

Definition 2.13 (regular scoring rule) Scoring rule S is *regular* relative to class $\mathcal{P}$, if it is real valued, except for the contingency of being $-\infty$ in case of $P \neq Q$.

If these two properties hold, then the divergence function associated with S is $d(P,Q) = S(Q,Q) - S(P,Q)$.

Theorem 2.1 *A regular scoring rule $S : \mathcal{P} \times \Omega \to \overline{\mathbb{R}}$ is (strictly) proper relative to class $\mathcal{P}$ if and only if there exists a (strictly) convex, real function G on $\mathcal{P}$ such that $S(P,\omega) = G(P) - \int G^*(P,\omega)dP(\omega) + G^*(P,\omega)$ for $P \in \mathcal{P}$ and $\omega \in \Omega$, where $G^*(P,\cdot) : \Omega \to \overline{\mathbb{R}}$ is a subtangent of G at the point $P \in \mathcal{P}$. (See [46].)*

The importance of the theorem is that if $|\Omega| < \infty$ and $G(P)$ is smooth, then $d(P,Q)$ is the Bregman-divergence associated with G, where the information measure or entropy function is $G(P) = \sup_{Q \in \mathcal{P}} S(Q,P)$.

Savage representation in [92] for categorical variables, and later, Schervish representation in [93][5] for dichotomous events ($|\Omega| = \{0,1\}$) are essential works in the discussion of proper scoring rules, or as Savage notes, to elicite personal probabilities. Here we mention a few scoring rules on categorical variables, which are not limited to categorical variables but have counterparts for continuous variables. The Brier and the logarithmic scores are used in the next section in order to compare the

[5] In [93], Schervish makes a witty distinction between the two aspects of evaluating forecasters: either based on who *has given* the best forecast in the *past*, or who *will give* the best in the *future*. The two perspectives are not identical under general circumstances.

evaluation methods in automobile insurance. Recall that for an individual, the conditional distribution of the number of accidents is Poisson, thus in our notations let $p_i := P(X = i|\Lambda = \hat{\lambda})$ $(i = 0, 1, 2, \ldots)$, i.e. the probabilities of certain claim numbers using the estimated $\hat{\lambda}$ as condition. Similarly, q_i $(i = 0, 1, 2, \ldots)$ is the same probability, but for the real λ frequency.

Suppose that measure P is an approximation of measure Q. The following score definition dates back to [13]. The q_i probabilities are only known in theory and they can be used in the simulations as presented later in section 2.6, supporting decision making.

Definition 2.14 (Brier score) Let the Brier score be defined as

$$S(P, Q) = 2\left(\sum_i p_i q_i\right) - \left(\sum_i p_i^2\right) - 1. \tag{2.11}$$

Due to the associated Bregman divergence $d(p, q) = \sum_i (p_i - q_i)^2$, the Brier score is sometimes referred to as quadratic score. The entropy function is $G(p) = \sum_{i \in |\Omega|} p_j^2 - 1$. For an analysis regarding precipitation forecasts which use Brier scores see [38]. For the purpose of calculating the score of estimation subsequently, change q_i values to a Dirac delta, depending on the number of claims caused. In other words, if the examined policyholder had i claims last year, then it implies the corresponding score to be

$$S(P, i) = 2\left(\sum_j p_j \delta_{ij}\right) - \left(\sum_j p_j^2\right) - 1 = 2p_i - \sum_j p_j^2 - 1. \tag{2.12}$$

Definition 2.15 (logarithmic score) The logarithmic score is defined as

$$S(P, Q) = \sum_i q_i \log p_i. \tag{2.13}$$

See [49]. In case of i accidents caused, $S(P, i) = \log p_i$. The associated Bregman divergence is the Kullback-Leibler divergence $d(p, q) = \sum_i q_i \log \frac{q_i}{p_i}$. The entropy function is the negative Shannon entropy $G(p) = \sum_{i \in |\Omega|} p_i \log p_i$.

See appendix A for scores excluded from this chapter. Further concepts formalised directly in terms of the predictive distributions, such as the continuous ranked probability score and its generalisation, the energy score are discussed in the subsequent chapters.

2.6 Simulation and results

Several portfolios have been simulated in `R` in order to (1) approximate the parameters of the gamma mixing distribution from a test population as described in section 2.3 (2) estimate each policyholder's claim frequency as in section 2.4 and (3) compare the method using scores from section 2.5. The simulation technique can be used for frequency evaluation in practice, provided that we have the historical observations. To support decision making related to the choice of the best evaluation technique, we construct a ranking of estimation methods using score measures, given different inputs described in section 2.4. The number of years spent in the system by a driver has a significant effect on the prioritisation of methods. This impact underlines the fact that the convergence to an individual's stationary distribution can be extremely slow, and a method most advantageous for one insured might be outperformed by another method for another insured. In the examples this relation will be explained as a function of the number of observation years. It is important to emphasise that the comparison methodology in general might be applied for other actuarial models as well. Note that the example simulations below are limited to Poisson distributed claim numbers with Gamma mixing distribution for frequencies.

We need a first (study) portfolio containing N insured individuals, which is used for the estimation of the α and β parameters of the negative binomial distribution. These are the policyholders' histories available in the insurer's database. Henceforth, we refer to this institution as *our company*. Taking advantage of the entire claim and bonus–malus history, we do the parameter evaluation with MM1 from section 2.3. In the next step, we generate the history of a second (evaluation) portfolio containing M policyholders, assuming that the distribution parameters are unchanged compared to the first portfolio.

After that, we estimate the claim frequency parameters of policyholders separately, based on the three estimation methods described in section 2.4, and compare them to the real λ parameters using the Brier and the logarithmic scores. The objective is to decide which of the three methods would give the best-fit results in certain cases. The method that results in the higher average score value means the better goodness-of-fit.

Subsequently, we apply a Monte Carlo type technique. This means that we generate the above mentioned two portfolios r times independently, but in each first (study) one preserving the α and β distribution inputs. After that, based on the approximated $\hat{\alpha}$ and $\hat{\beta}$, we estimate the λ parameters of the second portfolios. Each method gives one score number for each simulation, which is the average of scores calculated for individuals. At last we take the mean of mean scores, and the method resulting the higher score is considered to be the better one.

Algorithm 2.1 (Arató, Martinek)

step I | For $i = 1, \ldots, r$

step A | Generate portfolio $P_i^{(1)}$ consisting of N individuals, using parameters α, β. (Expected claim numbers and claim histories.)

step B | Calculate estimators $\hat{\alpha}^{(i)}, \hat{\beta}^{(i)}$ related to α, β from $P_i^{(1)}$.

step C | Generate portfolio $P_i^{(2)}$ containing M policies, with parameters α, β. It results in $\lambda_1^{(i)}, \ldots, \lambda_M^{(i)}$ frequencies.

step D | According to $\hat{\alpha}^{(i)}, \hat{\beta}^{(i)}$ and each estimation method ($k = 1, 2, 3$), calculate estimates $\hat{\lambda}_1^{(i),k}, \ldots, \hat{\lambda}_M^{(i),k}$.

step E | Assign score $S_j^{(i),k}$ to each individual j ($j = 1, \ldots, M$, $k = 1, 2, 3$ method), and calculate the mean value $\overline{S}^{(i),k} = \frac{1}{M}(S_1^{(i),k} + \ldots + S_M^{(i),k})$.

step II | In accordance with the above detailed notations, the ultimate score value regarding one estimation method is $S^{\text{meth } k} = \frac{\overline{S}^{(1),k} + \ldots + \overline{S}^{(r),k}}{r}$.

step III | Order $S^{\text{meth } k_1} < S^{\text{meth } k_2} < S^{\text{meth } k_3}$, the higher the score the better the model's performance.

As a reference line we write and plot the score results that compare the real frequencies with the real frequencies (see $S(Q, Q)$). These scores are time-invariant, contrary to the charts below, where small differences can be observed. The reason is that in the Monte Carlo simulations we also generated the λ parameters over and over. Simulation input parameters are:

1. Distribution parameters α and β. These can be considered as the true underlying parameters, unknown to the insurance firm.
2. N the number of policyholders in the first (study) portfolio, which is used to estimate the α and β parameters.
3. M the number of policyholders in the second (evaluation) portfolio. This contains individuals, whose λ parameters have to be evaluated.
4. The number of steps in years. This means the time elapsed in years, since the certain individual is insured by our company. Note that it implies the knowledge of claims history and bonus classification, it is therefore very important information which affects the goodness-of-fit of the estimation methods.

5. The number of years elapsed before entering our company. This affects *method 1* and *method 2*, because the Markov chain on the bonus classes converges slowly to the stationary distribution.
6. Transition rules of the examined country.
7. The number of simulated portfolios. As we approximate the scores via Monte Carlo type technique, it has to be appropriately large.

We simulate a portfolio containing $N = 80{,}000$ people $r = 50$ times, estimate α and β parameters, then estimate the λ parameters of $M = 20{,}000$ policyholders. After that we set the results of the three methods against the real frequencies using scores. For every simulation and country the Brier and Log scores are calculated, see Figure 2–4, Figure 2–5 and Figure 2–6. The points of the charts represent the average $S^{\text{meth } k}$ scores as a function of time (year steps). Certain year steps mean that we generated the second portfolios (the current one we are analysing) as we had information about policyholders $1, 2, 5, 10, 15$ and 20 years retroactively. Intermediate points are approximated linearly. Note that the standard deviation of the sample of the Brier scores in the Hungarian example is under 0.0009, and under 0.0016 in case of the Log scores.

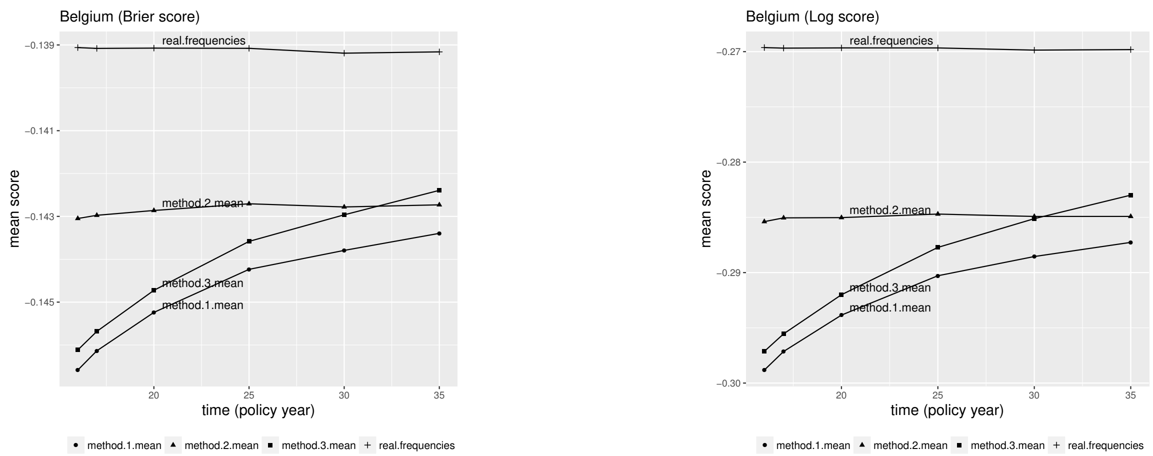


Figure 2–4: Brier and Log scores in the Belgian system. Mixing distribution parameters are $\alpha = 1.2$, $\beta = 14$.

In practice, there are different lengths of claim histories available for different policyholders. As a function of this length, we can decide that the parameters of a group of insureds will be evaluated using *method 2*, and the rest using *method 3*.

We let the random walks of policyholders in the systems run for 15 years on Figure 2–4, Figure 2–5 and Figure 2–6. Then they are assumed to be acquired by our company, which implies our observations to start at that point. In other words, year steps start at that time, when each driver has already spent some time randomly

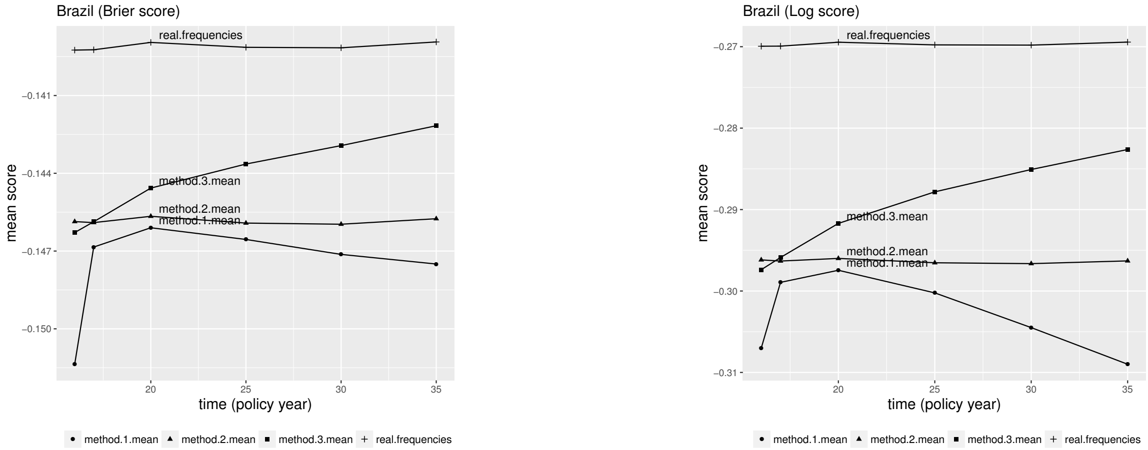


Figure 2–5: Brier and Log scores in the Brazilian system. Mixing distribution parameters are $\alpha = 1.2$, $\beta = 14$.

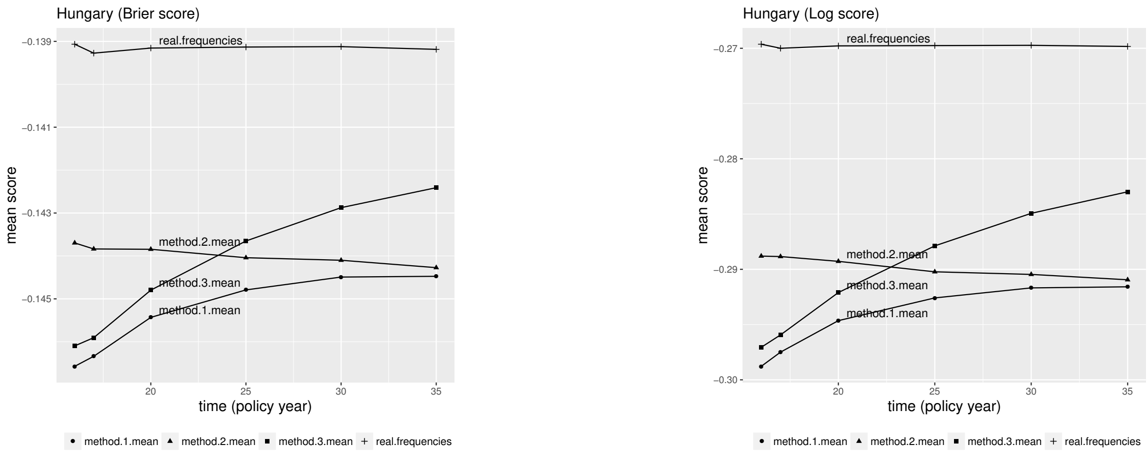


Figure 2–6: Brier and Log scores in the Hungarian system. Mixing distribution parameters are $\alpha = 1.2$, $\beta = 14$.

walking on the transition graph.

The examples clearly show the differentiation capability of systems containing more bonus–malus classes. The more classes the system has, the more years are needed for *method 3* to outperform *method 2*. This is due to the speed of convergence. This observation itself is not surprising, but the value of time until the exceedance is. The Bayesian type *method 1* consistently underperforms the other methods. However, if the claim history is largely deficient there can be no option to use 2 or 3.

The two types of scores have given almost identical results on the parameters tested (the difference is not significant). If *method x* is the best according to Brier scores, then it is the best according to Log scores as well.

2.7 Conclusion

In this chapter the principles of bonus–malus systems have been introduced, and the distributional assumptions regarding the claim numbers of the policyholders have been explained. We have assumed that the number X of claims caused in a year by an insured person is conditionally Poisson distributed given λ. The goal has been to evaluate the λ frequency parameters, which are the expected numbers of individual claims. Addressing the size of them is beyond the objectives of this chapter. Since the unconditional distribution is negative binomial, we can evaluate the shape and rate parameters based on the insurer's claim history from past years. The current portfolio may differ in parameters (study vs evaluation portfolios). The accurateness of MM and ML on portfolios of different sizes has been verified.

We have introduced 3 methods for frequency estimation, where the first one has been published first in [3] to our knowledge. The other two methods are known. Our aim was to decide which method is the most appropriate in certain circumstances, i.e. for the parameters given. Our decision has been made based on scores that measure the bias of two distributions. If *method x* results in $\hat{\lambda}_1, \ldots, \hat{\lambda}_M$ frequency parameters, and the real ones are $\lambda_1, \ldots, \lambda_M$, then *method x* is the best choice among its peers, if the average score $S^{\text{meth } k}$ is greater than in other cases. The discussion includes a Monte Carlo type algorithm, which can be used in practice to make decisions.

Furthermore, our method is a technique suitable to compare bonus–malus systems in the following sense: As a function of years, the longer the *method 2* is better than the others, the more informative the system is, as we expect more accurate evaluation of claims frequencies using the past year's average claim numbers in different classes. Supposing that the real parameters are α and β, in the Brazilian system, *method 3* based on claims history becomes the most appropriate in the second year, while in the Hungarian system it needs 7-8, and in the Belgian 16-17 years.

The ranking method can be applied under other distributional and model constraints as well.

3 An Approach to Merit Rating by Means of Autoregressive Sequences

3.1 Introduction

Numerous original research papers are dedicated to the construction of optimal bonus–malus systems, which primarily relate to a set of questions rather than to a specifically defined problem. The majority of research focusses on the optimal adjustment of relativities. One essential article from the earlier period of optimisation analysis is [80], which minimises the squared difference in expectation between the asymptotical expected claim and the premium, given the stationary distribution of the random walk on premium classes (comparable to corollary 3.7). Under financial stability constraint, further progress is made by [26], by analysing the quadratic as well as an exponential loss function. [106] is an extension of the authors' earlier works that combines the claim frequencies with the severity (size) of the claim under several distributional constraints. Furthermore, the model proposed by these authors takes both individual a priori and a posteriori information into account. Another more recent paper [101] creates optimal relativities in accordance with a proposed objective function and applies varying transition rules. [48] makes a distinction between the larger, more costly claims and the smaller, less costly losses and proposes a statistical model incorporating a bivariate distribution, which demarcates losses relative to a certain threshold. This approach enables insurance companies to form the premiums such that policyholders with fewer, less significant claims are penalised to a lesser extent than policyholders with larger, more costly claims. This paper provides numerical simulations that are performed on the basis of real data from the Macquarie University, Australia [54].

The correlation between covariates such as age and gender, for instance, has been investigated in several papers. As another illustration of utilising real portfolio data, [63] validates Taylor's Bayesian simulation model using Taiwanese claim history.

Furthermore, [86] have acquired a Belgian portfolio for their analysis.

In contrast to the enumerated papers, instead of analysing the more optimal finite set of relativities, the present chapter proposes a fundamentally different set of transition rules from one premium class to another. To summarise, the purpose of the present chapter is (1) to introduce a new scheme which is structurally different from those in use, (2) to evaluate its metrics of higher relevance, and (3) to put it into context by comparing it with existing models. Note that the description of existing models will be limited to what is relevant for comparison purposes only; detailed elaboration on these models is beyond the scope of this chapter.

Section 3.2 describes a first order autoregressive model built on premiums and individual claim history. In contrast to the models with a finite state space, the number of possible relativity levels of this model is not bounded. Claim frequencies are supposed to follow a stationary Markov chain that reflects the successive links among claims of a policyholder throughout the years. The section also includes analytical formulas to calculate some of the relevant measures of a bonus–malus system, such as elasticity, coefficient of variation, (modified) relative stationary average premium level and financial equilibrium. Definitions in the general case with stationary Markov chain frequencies require the reinterpretation of original measure concepts.

In previous decades insurance institutions received more freedom from central regulatory bodies in the selection of their experience rating systems. European countries in section 3.3 are authorised to determine their premium relativities. Certainly, in practice schemes can vary widely in terms of the rules applied and the actual setting of relativities. The comparison of the new model to the existing schemes in Belgium, Hungary and the Netherlands illustrates the discrepancies in the metrics listed in section 3.3. Section 3.4 concludes the chapter. This part of the dissertation is based on [73].

3.2 Recursive premium model

3.2.1 Insurance claims processes

Let $Y_1, Y_2, \ldots$ be a homogeneous stationary Markov chain on a finite state space of *base frequencies* $\{\zeta_1, \zeta_2, \ldots, \zeta_p\}$ with non-negative real numbers $0 \leq \zeta_1 < \ldots < \zeta_p < \infty$. A fixed policyholder has an individual $\lambda > 0$ constant.

Definition 3.1 (frequency process) Let $\Lambda_t = \lambda \cdot Y_t$ be the *frequency process* of the insured.

The process is defined to be more general than in chapter 2. This parameterisation ensures that the classical definitions such as elasticity can be extended in a meaningful manner. Let $X_1, X_2, \ldots$ denote the claim number process of the policyholder. Suppose that given a fixed sequence $\Lambda_1 = \lambda_1, \Lambda_2 = \lambda_2, \ldots$, $X_1, X_2, \ldots$ are independent Poisson distributed random variables with parameters $\lambda_1, \lambda_2, \ldots$.

Suppose that the transition probability matrix $\mathbf{P}$ of the stationary Markov chain $\{Y_t\}$ corresponds to an ergodic Markov chain and hence, $\mathbf{P}$ has no eigenvalues with modulus greater than 1. $\mathbf{P} \in \mathbb{R}^{p \times p}$ and $(P)_{i,j} = P(Y_{t+1} = \zeta_j | Y_t = \zeta_i)$. Furthermore, let $\pi = \begin{pmatrix} \pi_1 & \ldots & \pi_p \end{pmatrix}^\top$ denote the stationary distribution of process Y_t, where $\pi_k = P(Y_t = \zeta_k)$.

Claim numbers are counted for consecutive and equally long periods of time, i.e. X_t stands for the tth year claim. Suppose from now on that periods represent years. Claim numbers should not necessarily be governed by Poisson distribution, however, using other assumptions may result in losing the analytical properties of the formulas. Possible modifications to the Poisson distribution with practical relevance are proposed in [112]. The zero-inflated Poisson distribution resulting from real-life portfolio observations is motivated by the number of zero claims, which is lower than the original Poisson would imply. Our model incorporates this assumption as well as a special case. In fact, it is not necessary for the governing claim number distribution to be discrete, placing the significance on the total size of claims rather than the count.

Example 3.1 (Markov chains) If $Y_t \equiv 1$, then X_t random variables are independent and Poisson distributed with parameter λ.

Example 3.2 (Markov chains) Let p be a fixed positive integer, $0 \leq \zeta_1 < \ldots < \zeta_p$ real numbers, and let the one-step transition probabilities be

$$\begin{aligned} P(Y_{t+1} = \zeta_{k+1} | Y_t = \zeta_k) &= \nu, \; k = 1, \ldots, p-1 \\ P(Y_{t+1} = \zeta_{k-1} | Y_t = \zeta_k) &= \mu, \; k = 2, \ldots, p \\ P(Y_{t+1} = \zeta_k | Y_t = \zeta_k) &= 1 - \mu - \nu, \; k = 2, \ldots, p-1 \\ P(Y_{t+1} = \zeta_k | Y_t = \zeta_k) &= 1 - \nu, \; k = 1 \\ P(Y_{t+1} = \zeta_k | Y_t = \zeta_k) &= 1 - \mu, \; k = p \end{aligned} \tag{3.1}$$

with real numbers $0 < \mu < 1$, $0 < \nu < 1$, $\mu + \nu \leq 1$. In other words, the chance of migrating to a frequency one level higher is ν, whilst to a frequency one level lower is μ with the exception in the lowest and highest ζ_1 and ζ_p. An important special case is when $p = 1$, coinciding with example 3.1.

Proposition 3.1 *The stationary distribution* π *of the Markov chain described by equation 3.1 is* $\pi_k = \left(\frac{\nu}{\mu}\right)^{k-1} \cdot \frac{1-\frac{\nu}{\mu}}{1-\left(\frac{\nu}{\mu}\right)^p}$, $k = 1, \ldots, p$ *if* $\nu \neq \mu$ *and* $\pi_k = \frac{1}{p}$ *otherwise.*

Proof. The stationary distribution π is the right eigenvector corresponding to eigenvalue 1 of the transition probability matrix defined by equation 3.1. Hence,

$$\begin{aligned} \pi_1 &= (1-\nu)\pi_1 + \mu\pi_2 \\ \pi_k &= \nu\pi_{k-1} + (1-\mu-\nu)\pi_k + \mu\pi_{k+1}, \;\; k = 2, \ldots, p-1 \\ \pi_p &= (1-\mu)\pi_p + \nu\pi_{p-1}, \end{aligned} \tag{3.2}$$

implying that $\pi_k = \left(\frac{\nu}{\mu}\right)^{k-1} \pi_1$ for $k = 1, \ldots, p$. Thus, $\sum_{k=0}^{p-1} \left(\frac{\nu}{\mu}\right)^k = \frac{1}{\pi_1}$, which proves the proposition. ■

As a consequence of the assumptions above, X_t is a stationary process and its moments can be determined for any positive integer l, see the next subsection.

3.2.2 Autoregressive construction of merit rating

Let $r_1 > 0$ be the initial value of the premium level, a constant assigned to the premium payment of any policyholder in the first year or period insured.

Let each subsequent premium be determined by the previous period's premium level r_t and claim number X_t, as follows:

$$r_{t+1} = r_t \cdot (1-c) + f(X_t) \cdot d, \tag{3.3}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a function of the number of claims in year t. Furthermore, let $0 < c < 1$ and $d > 0$ be real numbers. Hence, r_t is a stochastically recursive process, more precisely a first order autoregressive process. It is non-Gaussian and the role of the white noise is exchanged by a function of a sequence of Poisson(λ_t) variables. The general process is therefore not a white noise but a stationary process, as the sequence of noise does not consist of independent random variables, except for $p = 1$. The premium in time T can be expressed as

$$r_T = r_1(1-c)^{T-1} + \sum_{k=1}^{T-1} f(X_k) \cdot d \cdot (1-c)^{T-1-k}, \tag{3.4}$$

which has a tendency to diminish the initial premium level and older claims. The larger the constant c, the faster the initial premium level and older claims are lost in the process. Observe that the rule of transition from one premium state to another is relatively simple in practice, provided that function $f(x)$ is simple. Each subsequent price stems from the previous one by simultaneously decreasing it with a fixed percent and increasing it with a function of the claims observed. Parameter d can in fact be merged into function $f(x)$ without loss of generality, however, throughout this chapter d is indicated separately.

Definition 3.2 (stationary distribution) Let Q_{r^*} denote the stationary distribution of the process if it exists, and r^* a random variable with distribution Q_{r^*}.

Theorem 3.2 (Arató, Martinek) *Let $\xi_n := d \cdot f(X_n)$ be a stationary process and $E\,|f(X_n)| < \infty$ (with an arbitrary function $f : \mathbb{R} \to \mathbb{R}$ in equation 3.3). Then equation $r_n = (1-c)r_{n-1} + \xi_{n-1}$ has a stationary solution.*

Theorem 3.2 is a consequence of [89] in the case of independent X_n random variables. Equation 3.3 has a stationary solution if and only if the polynomial $p(z) = 1 - (1-c)z$ has no root of absolute value greater than or equal to 1, i.e. $|1-c| < 1$, which is true by definition in our case. The original statement in the reference requires that $E(f(X_t)) = 0$, however, this constraint can be omitted here. If $\tilde{f}(X_t) = f(X_t) - E(f(X_t))$ and $\tilde{r}_{t+1} = \tilde{r}_t(1-c) + d\tilde{f(X_t)}$, then $\tilde{r}_t$ obeys the strict stationarity, and $\lim\limits_{t\to\infty} r_t = \lim\limits_{t\to\infty} \tilde{r}_t + d \sum\limits_{k=1}^{T-1} E(f(X_k))(1-c)^{T-1-k} = \lim\limits_{t\to\infty} \tilde{r}_t + \frac{d}{c}E(f(X_1))$. Here we provide a direct proof of the above theorem for the general stationary process. It must be emphasised that the stationarity of r_t and X_t should be addressed separately, as the second one is defined to be stationary, whilst the first one begins in a fixed constant state.

Proof.(theorem 3.2) Fix a real c_0 for which $|1-c| < |1-c_0| < 1$ and an arbitrary integer n. Let $A_m := \{|(1-c)^m \xi_{n-m}| \geq (1-c_0)^m\}$.

$$P(A_m) \leq \frac{E\,|(1-c)^m \xi_{n-m}|}{(1-c_0)^m} = \frac{(1-c)^m}{(1-c_0)^m} E\,|\xi_{n-m}|\,, \tag{3.5}$$

where the inequality follows from Markov's inequality and $\kappa := E|\xi_{n-m}|$ is a finite constant. Thus, according to the Borel–Cantelli lemma,

$$\sum_{m=0}^{\infty} P(A_m) \leq \sum_{m=0}^{\infty} \left(\frac{1-c}{1-c_0}\right)^m \cdot \kappa < \infty \tag{3.6}$$

implies that the probability that infinitely many events A_m occur is zero. Hence,

$$\limsup_{m\to\infty} |(1-c)^m \xi_{n-m}|^{1/m} \leq (1-c_0) < 1. \tag{3.7}$$

This proves that r_n is almost surely convergent. Furthermore, $\sum\limits_{k=1}^{\infty} (1-c)^k \xi_{n-k} =$

$(1-c)\sum_{k=0}^{\infty}(1-c)^k\xi_{n-1-k}+\xi_n$ ensures that r_n is stationary. ■

Corollary 3.3 *Suppose that $f(x) = a_n x^n + \ldots + a_1 x + a_0$ is an arbitrary polynomial, and let equation 3.3 define the premium level structure. If these conditions apply, the stationary distribution of the premium level exists. Observe that the X_t claim sequence is a stationary process, since the frequency process is stationary as well.*

Proposition 3.4 *Suppose that the stationary distribution exists and its expectation is finite. The expected value of the stationary state is $E(r^*) = \frac{d}{c}E(f(X_1))$. As a special case, if $f(x) = x$, the expectation is $\frac{d}{c}E(\lambda_t)$.*

Proof. The existence of the stationary distribution entails that $r^* \overset{d}{=} r^*(1-c) + f(X_1)d$. Taking expectation results in $E(r^*) = (1-c)E(r^*) + E(f(X_1))d$, which proves the proposition after rearranging the equation. ■

From now on let $f(x) = a_n x^n + \ldots + a_1 x + a_0$ be a polynomial of degree n. The properties of such a premium system are subject to scrutiny in the following sections. The non-zero constant coefficient in polynomial f is not excluded from the model, however, a_0 being positive implies that the premium cannot reach a level below $a_0 d/c$, forming a lower bound of r_t. From equation 3.3 it can be seen that if for any period t, $r_t \cdot c < a_0 \cdot d$, then the subsequent premium will strictly be higher than r_t, after a claim-free period t. Eventually, a non-zero a_0 is a tradeoff which allows for a premium's strictly positive lower limit and implies that the premium monotonicity following a claim-free year fails at the same time. If a sufficiently low constant element a_0 is selected, the rebound is not significant and a minimal premium value is guaranteed.

A claim-free period will clearly mean a deduction in the premium level $r_{t+1} = r_t \cdot (1-c) + a_0 d + 0$ (except for $r_t c < a_0 d$). One claim in period t results in a subsequent rating of $r_{t+1} = r_t \cdot (1-c) + d \cdot (a_n + \ldots + a_1 + a_0)$. The distinctive effect of the polynomial structure is reflected only in the case of two or more claims, giving more weight to the coefficients of higher degrees of x (i.e. the powers of a claim number larger than 1 obviously do not weigh $a_0, \ldots, a_n$ evenly).

It is inevitable to make a distinction between the frequency states of the policyholders. As noted in subsection 3.2.1, each base state ζ_i of frequencies is multiplied by a personal constant λ. Let the premium process be labeled with $r_n^{(\lambda)}$. Furthermore, $r^{*(\lambda)}$ denotes an element from the stationary state of $r_n^{(\lambda)}$.

Theorem 3.5 (Arató, Martinek) *Suppose that $f(x) = a_n x^n + \ldots + a_1 x + a_0$ and that Λ_t is the frequency process as defined in subsection 3.2.1. Let $\underline{a} := \begin{pmatrix} a_0 & a_1 & \ldots & a_n \end{pmatrix}^\top$. Let $B_{(n)}$ be an $(n+1) \times (n+1)$ matrix with elements b_{ij}, equal to the coefficient* *corresponding to the j^{th} degree term in the Bell polynomial of degree i, where indices are $i,j = 0,\dots,n$. Furthermore, let $Z_{(n)} \in \mathbb{R}^{(n+1)\times(p+1)}$, $(Z_{(n)})_{i,j} = \begin{cases} \zeta_j^i, & i \neq 0 \cap j \neq 0 \\ 1, & i = j = 0 \\ 0, & otherwise \end{cases}$ $(i = 0,\dots,n$ and $j = 1,\dots,p)$. Lastly, $\pi' = \begin{pmatrix} 1 \\ \pi \end{pmatrix}$, where π is the stationary distribution of Λ_t and $diag(\lambda^0, \lambda^1, \lambda^2, \dots, \lambda^n)$ is a matrix with λ^is in the diagonal and 0 elsewhere. Then*

$$E(f(X_t)) = \underline{a}^\top B_{(n)} diag(\lambda^0, \lambda^1, \dots, \lambda^n) Z_{(n)} \pi'. \tag{3.8}$$

Proof. The moment-generating function of random variable $X \sim Poisson(\lambda)$ is $M_X(t) = e^{\lambda(e^t-1)}$. Dobiński's formula states that the Bell polynomials have similar exponential generating function $\sum\limits_{k=0}^{\infty} \frac{B_k(x)}{k!} t^k = e^{x(e^t-1)}$. Hence, the n^{th} moment of the Poisson distribution can be expressed with the Bell function of degree n: $E(X^n) = M_X^{(n)}(t)|_{t=0} = B_n(\lambda)$. Let $B_n(\lambda) = b_{n,n}\lambda^n + b_{n,n-1}\lambda^{n-1} + \ldots + b_{n,1}\lambda$ denote the Bell polynomial of degree n, for which $b_{n,n} = 1\ \forall n$, moreover, $b_{n,0} = 0$ for $n > 0$. Thus, for $k \geq 1$ using the law of total expectation

$$E(X_t^k) = \sum_{i=1}^{p} \pi_i E(X_t^k | Y_t = \zeta_i) = \sum_{i=1}^{p} \pi_i B_k(\lambda \cdot \zeta_i) = \sum_{i=1}^{p} \pi_i \sum_{l=1}^{k} b_{k,l} \cdot \lambda^l \cdot \zeta_i^l, \tag{3.9}$$

$$Ef(X_t) = a_0 + \sum_{k=1}^{n} a_k \sum_{i=1}^{p} \pi_i \sum_{l=1}^{k} b_{k,l} \cdot \lambda^l \cdot \zeta_i^l = \underline{a}^\top B_{(n)} diag(\lambda^0, \lambda^1, \dots, \lambda^n) Z_{(n)} \pi'. \tag{3.10}$$

■

Corollary 3.6 *As a special case, if $\Lambda_t \equiv \lambda$ and $\underline{\lambda}_{(n)} = \begin{pmatrix} 1 & \lambda & \lambda^2 & \dots & \lambda^n \end{pmatrix}^\top$, then $E(f(X)) = \underline{a}^\top B_{(n)} \underline{\lambda}_{(n)}$.*

Observe that the i^{th} row $(i = 0, 1, \dots, n)$ of $B_{(n)}$ consists of the coefficients of $B_i(\lambda)$, i.e. $b_{i,0}$, $b_{i,1}, \dots, b_{i,i}$ and $b_{i,j} = 0$ for $j > i$. In other words, $B_{(n)}$ is a lower triangular with Bell coefficients in each row. All the values are 0 in the first column, except for the first row:

$$B_{(n)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 3 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 6 & 7 & 1 & 0 & 0 & \dots \\ & & & \vdots & & & & \ddots \end{pmatrix}. \tag{3.11}$$

Corollary 3.7 *For a stationary Markov chain Λ_t, the squared difference between the stationary premium level and λ can be expressed as*

$$\begin{aligned} (E(r^* - \Lambda_t))^2 &= \left(\frac{d}{c}E(f(X_t)) - E(\Lambda_t)\right)^2 \\ &= \left(\frac{d}{c}\underline{a}^\top B_{(n)} diag(\lambda^0, \lambda^1, \dots, \lambda^n) Z_{(n)}\pi' - \underline{e}_1^\top diag(\lambda^0, \lambda^1, \dots, \lambda^n) Z_{(n)}\pi'\right)^2 \\ &= \left(\left(\frac{d}{c}\underline{a}^\top B_{(n)} - \underline{e}_1^\top\right) diag(\lambda^0, \lambda^1, \dots, \lambda^n) Z_{(n)}\pi'\right)^2, \end{aligned} \tag{3.12}$$

where $\underline{e}_1^\top = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \end{pmatrix}$. *For* $\Lambda_t \equiv \lambda$, $(E(r^*) - \lambda)^2 = \left(\left(\frac{d}{c}\underline{a}^\top B_{(n)} - \underline{e}_1^\top\right) \underline{\lambda}_{(n)}\right)^2$.

In a special case, $\underline{a}^\top = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \end{pmatrix}$ and $d = c$ implies that the squared difference $(E(r^*) - E(\Lambda_t))^2$ is zero for any stationary Markov chain Λ_t. This special case distributes premiums fairly among policyholders over the long term in a democratic manner. However, the compromise required in order to achieve this fitting results in an excessively high volatility in price, which curtails the practical benefits of such a parameterisation.

3.2.3 Elasticity

In the long run the expectation of a system is that the premium paid is a monotonically increasing function of the claim frequency (constant) λ. Ideally, this relation is linear. In other words, any fraction of increment $\frac{\partial E(r^{*(\lambda)})}{E(r^{*(\lambda)})}$ in the mean value of stationary premium is associated with an identical increment in the expected number of claims $\frac{\partial \lambda}{\lambda}$. However, the stationary Markov chain model of claims frequencies encompasses a broader concept of stationary premiums, requiring the clarification of the derivatives mentioned in the original concept introduced by [67]. This concept expresses the asymptotical premium as dependent on the change in claim frequency, i.e. the premium gets more expensive in a fair proportion as the average claim number increases. Here we introduce a slight generalisation of the original definition, suitable for Λ_t claim frequencies. It is a generalisation in the sense that it uses a function of the constant multiplier in a stationary Markov chain, allowing for a variable claim frequency process instead of fixed λ parameters. In the $\Lambda_t \equiv \lambda$ case it is equivalent to the definition of Loimaranta.

Definition 3.3 (elasticity) Let the claim frequency process be as defined in subsection 3.2.1, a stationary Markov chain on state space $\{\zeta_1, \ldots, \zeta_p\}$, multiplied by $\lambda > 0$. The modified elasticity or Loimaranta efficiency of the stationary premium is

$$\eta(\lambda) = \frac{\frac{\partial E(r^{*(\lambda)})}{\partial \lambda}}{\frac{E(r^{*(\lambda)})}{\lambda}}.$$

Proposition 3.8 *Let Z be an $(n+1) \times (p+1)$ matrix as defined in theorem 3.5. Matrix $diag(\lambda^0, \lambda^1, \ldots, \lambda^n) \in \mathbb{R}^{(n+1)\times(n+1)}$ contains λ^is in the diagonal and 0 elsewhere, whilst $D_{(n)}$ is $diag(0, 1, \ldots, n)$. Furthermore, $\pi' = \begin{pmatrix} 1 \\ \pi \end{pmatrix}$. The elasticity is*

$$\eta(\lambda) = \frac{\underline{a}^\top B_{(n)} D_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'}{\underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'}. \tag{3.13}$$

Proof.

$$\begin{aligned}
\eta(\lambda) &= \frac{\frac{\partial E(r^{*(\lambda)})}{\partial \lambda}}{\frac{E(r^{*(\lambda)})}{\lambda}} = \frac{\frac{\partial \frac{d}{c} E(f(X_t))}{\partial \lambda}}{\frac{\frac{d}{c} E(f(X_t))}{\lambda}} = \frac{\frac{\partial \underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'}{\partial \lambda}}{\frac{\underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'}{\lambda}} \\
&= \frac{\underline{a}^\top B_{(n)} \left(\frac{1}{\lambda} D_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)}\right) \underline{\pi}'}{\frac{\underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \underline{\pi}'}{\lambda}} \\
&= \frac{\underline{a}^\top B_{(n)} D_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'}{\underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi'},
\end{aligned} \tag{3.14}$$

where the numerator uses

$$\frac{\partial}{\partial \lambda} \underline{a}^\top B_{(n)} diag(\lambda^0, \ldots, \lambda^n) Z_{(n)} \pi' = \sum_{k=1}^{n} a_k \sum_{i=1}^{p} \pi_i \sum_{l=1}^{k} b_{k,l} \cdot (l-1) \lambda^{l-1} \cdot \zeta_i^l,$$

where instead of $diag(\lambda^0, \ldots, \lambda^n)$, the multiplier in the formula then becomes $\frac{1}{\lambda} \cdot diag(0, \ldots, n) \cdot diag(\lambda^0, \ldots, \lambda^n)$. ■

Corollary 3.9 *If $\Lambda_t \equiv \lambda > 0$, then $\eta(\lambda) = \frac{\underline{a}^\top B_{(n)} D_{(n)} \underline{\lambda}}{\underline{a}^\top B_{(n)} \underline{\lambda}}$.*

According to proposition 3.8, elasticity can only be perfect, i.e. $\eta \equiv 1$ in the case of $n = 1$ (if $a_i \neq 0$ only for $i = 1$ and 0 otherwise). For polynomials of higher degree the system loses its perfection.

3.2.4 Coefficient of variation

A commonly analysed property of a merit rating system is the ratio of the standard deviation to the expected value of the stationary distribution. A major purpose of insurance is the financial stability of policyholders through the mitigation of large losses and the maintenance of a relatively flat price; it is therefore reasonable to expect an insurance scheme to lack substantial jumps in premiums. In practice, the stationary state can be beyond the foreseeable future from the perspective of the lifetime of insurance policies, as [6] points out, or as we have seen in section 2.6. This challenges the stationary point of view by introducing age-correction in ordinary bonus–malus systems.

Definition 3.4 (coefficient of variation) Let $s = \frac{\sqrt{Var(r^*)}}{E(r^*)}$ denote the *coefficient of variation* of the stationary process. (Occasionally another definition is used, in which $r^*_{t+1} - r^*_t$ is the *change* in premium level supposing the stationary regime of the process. This second definition is $\frac{\sqrt{E(r^*_{t+1}-r^*_t)^2}}{E(r^*)}$. In the dissertation we will use the former, as it is more commonly used in the literature.)

Definition 3.4 generally holds for the stationary claim frequency process if expressed as a function of stationary premium. However, for the comparison between risk classes given a fixed base stationary Markov chain of claim frequencies (on state space $\{\zeta_1, \zeta_2, \ldots\}$), the coefficient of variation can be expressed as a function of parameter λ.

Theorem 3.10 (Arató, Martinek) *Let $\sigma_f(m) := Cov(f(X_n), f(X_{n+m}))$ denote the covariance between $f(X_n)$ and $f(X_{n+m})$ for any positive integer n. Suppose that the number of claims in a period has a finite variance, $Var(f(X_1)) < \infty$. The coefficient of variation is equal to*

$$\sqrt{\frac{c}{2-c}} \cdot \frac{\sqrt{\sigma_f(0) + 2\sum\limits_{k=1}^{\infty} \sigma_f(k) \cdot (1-c)^k}}{E(f(X_1))}. \tag{3.15}$$

Proof.

Assuming the stationary premium regime r_n and the abbreviation $\xi_n := d \cdot f(X_n)$,

$$
\begin{aligned}
Var(r_n) &= Cov\left(\sum_{k=0}^{\infty}(1-c)^k \cdot \xi_{n-k}, \sum_{l=0}^{\infty}(1-c)^l \cdot \xi_{n-l}\right) \\
&= d^2 \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(1-c)^{k+l} \cdot \sigma_f(k-l) \\
&= d^2 \left(\sigma_f(0) \cdot \frac{1}{1-(1-c)^2} + \sigma_f(1) \cdot \frac{2(1-c)}{1-(1-c)^2} + \sigma_f(2) \cdot \frac{2(1-c)^2}{1-(1-c)^2} + \dots\right) \\
&= d^2 \frac{1}{1-(1-c)^2} \cdot \sigma_f(0) + d^2 \frac{2}{1-(1-c)^2} \sum_{k=1}^{\infty} \sigma_f(k) \cdot (1-c)^k \\
&= d^2 \frac{1}{1-(1-c)^2} \cdot \left(\sigma_f(0) + 2 \cdot \sum_{k=1}^{\infty} \sigma_f(k) \cdot (1-c)^k\right).
\end{aligned}
\tag{3.16}
$$

With proposition 3.4 the coefficient of variation is equal to

$$
\frac{\sqrt{\frac{1}{1-(1-c)^2}}}{\frac{1}{c}} \cdot \frac{d\sqrt{\sigma_f(0) + 2 \cdot \sum_{k=1}^{\infty} \sigma_f(k) \cdot (1-c)^k}}{E\xi_1},
$$

proving the theorem after rearranging the first fraction and removing parameter d. ■

Corollary 3.11 *In the special case if $\Lambda_t \equiv \lambda$, the coefficient of variation is $\sqrt{\frac{c}{2-c}} \cdot \frac{\sqrt{Var(f(X_1))}}{E(f(X_1))}$.*

The connection of the coefficient of variation to the constant c reflects a tradeoff. The first multiplying factor is monotonically increasing as a function of c, resulting in a larger coefficient. However, if the value of c is low, the convergence to the stationary distribution becomes slow.

Proposition 3.12 *The autocovariance of the stationary process $f(X_t)$ is*

$$
\sigma_f(m) = \sum_{i=1}^{p}\sum_{j=1}^{p}\left(\underline{a}^\top B_{(n)}\underline{\zeta_i}_{(n)}\right) \cdot \left(\underline{a}^\top B_{(n)}\underline{\zeta_j}_{(n)}\right) \cdot \pi_i(\mathbf{P}^m)_{i,j} - \left(\underline{a}^\top B_{(n)} diag(\lambda^0, \dots, \lambda^n) Z_{(n)} \pi'\right)^2,
\tag{3.17}
$$

for $m > 0$, where $\underline{\zeta_i}_{(n)} := \begin{pmatrix} 1 & \zeta_i & \zeta_i^2 & \dots & \zeta_i^n \end{pmatrix}^\top$ and $\mathbf{P}$ is the transition probability matrix of process Λ_t.

Proof. $\sigma_f(m) = Cov(f(X_k), f(X_{k+m})) = E\left(f(X_k)f(X_{k+m})\right) - E^2(f(X_1))$, where the second expected value is already known, hence, the scrutiny of the first one remains. Exploit that for $m > 0$, terms $f(X_k)$ and $f(X_{k+m})$ are conditionally independent given $\{\Lambda_k, \Lambda_{k+m}\}$. Thus, we may write that

$$\begin{aligned} E\left(f(X_k)f(X_{k+m})\right) &= E\left(E\left(f(X_k)f(X_{k+m})|\Lambda_k, \Lambda_{k+m}\right)\right) \\ &= E\left(E\left(f(X_k)|\Lambda_k\right) \cdot E\left(f(X_{k+m})|\Lambda_{k+m}\right)\right). \end{aligned} \tag{3.18}$$

Observe that $E(f(X_k)|\Lambda_k) = \underline{a}^\top B_{(n)}\underline{\zeta_i}_{(n)}$. Furthermore, the probability that the stationary Markov chain shifts from state ζ_i to state ζ_j in m steps can be expressed as the (i,j) element of the mth power of the transition probability matrix, $(\mathbf{P}^m)_{i,j}$, proving the proposition. ■

Proposition 3.13 *Let $J_k \in \mathbb{R}^{(n+1)\times(n+1)}$ be a matrix where its elements are 1 if the indices satisfy $i + j = k - 1$ $(i, j = 0, 1, \ldots, n)$ and 0 otherwise, $k = 1, \ldots, 2n + 1$, $\underline{a} = (a_0\ a_1\ \ldots\ a_n)^\top$ is the array of polynomial coefficients in $f(x)$ and $B_{(n)}$ is the matrix of Bell coefficients as in equation 3.11. Furthermore, let $\underline{e}_i \in \mathbb{R}^{(2n+1)\times 1}$ be a vector with all zero elements, except the ith term is equal to 1, $i = 0, 1, \ldots, 2n$. Then*

$$E(f(X_t)^2) = \underline{a}^\top \left(\sum_{k=1}^{2n+1} J_k \cdot \left(\underline{e}_{k-1} B_{(2n)} diag(\lambda^0, \ldots, \lambda^{2n}) Z_{(2n)} \pi'\right)\right) \underline{a}. \tag{3.19}$$

Proof. Let $\underline{X}_t$ denote $(1\ X_t\ \ldots\ X_t^n)^\top$.

$$\begin{aligned} E(f(X_t)^2) &= E\left((\underline{a}^\top \underline{X}_t)^2\right) \\ &= E\left(\sum_{k=1}^{2n+1} \underline{a}^\top J_k \underline{a} \cdot X_t^{k-1}\right) \\ &= \sum_{k=1}^{2n+1} \underline{a}^\top J_k \underline{a} \cdot E\left(X_t^{k-1}\right) \\ &= \sum_{k=1}^{2n+1} \underline{a}^\top J_k \underline{a} \cdot \left(\underline{e}_{k-1} B_{(2n)} diag(\lambda^0, \ldots, \lambda^{2n}) Z_{(2n)} \pi'\right) \\ &= \underline{a}^\top \left(\sum_{k=1}^{2n+1} J_k \cdot \left(\underline{e}_{k-1} B_{(2n)} diag(\lambda^0, \ldots, \lambda^{2n}) Z_{(2n)} \pi'\right)\right) \underline{a}. \end{aligned} \tag{3.20}$$

The elements of matrix $J_k \in \mathbb{R}^{(n+1)\times(n+1)}$ are 1 if the index satisfies $i + j = k - 1$ $(i, j = 0, 1, \ldots, n)$ and 0 otherwise, $k = 1, \ldots, 2n + 1$. Furthermore, we used theorem 3.5 for expressing $E(X^{k-1})$. ■

Corollary 3.14 *Let $M_{(n)} \in \mathbb{R}^{(n+1)\times(n+1)}$ be defined as $\sum_{k=1}^{2n+1} J_k \cdot B_{k-1}(\lambda)$ and let $B_{k-1}(\lambda)$ be the Bell polynomial of degree $k - 1$. If $\Lambda_t \equiv \lambda$, then $E(f(X_t)^2) = \underline{a}^\top M_{(n)}(\lambda)\underline{a}$ and $Var(f(X))$ can be expressed as $\underline{a}^\top \left(M_{(n)}(\lambda) - B_{(n)}\underline{\lambda}\underline{\lambda}^\top B_{(n)}^\top\right) \underline{a}$.*

The premium construction has a disadvantage when λ is close to zero, provided that $a_0 = 0$. $Var(f(X))$ is a polynomial of λ with a non-zero coefficient corresponding to λ^1, whilst in the case of $E^2(f(X))$ the term of the lowest power with positive coefficient is λ^2. Thus, as λ tends to zero, the coefficient of variation tends to infinity.

On the contrary, if $a_0 \neq 0$, then the coefficient of variation tends to 0 as $\lambda \to 0$. Indeed,

$$s^2 = \frac{c}{2-c}\left(\frac{E(f(X)^2)}{E^2(f(X))} - 1\right) = \frac{c}{2-c}\left(\frac{\underline{a}^\top \left(\sum_{k=1}^{2n+1} J_k \cdot B_{k-1}(\lambda)\right)\underline{a}}{\underline{a}^\top \left(B_{(n)}\underline{\lambda}\underline{\lambda}^\top B_{(n)}^\top\right)\underline{a}} - 1\right) = \tag{3.21}$$
$$= \frac{c}{2-c}\left(\frac{a_0^2 + \lambda \cdot \ldots}{a_0^2 + \lambda \cdot \ldots} - 1\right) \overset{\lambda \to 0}{\to} 0.$$

Corollary 3.14 also implies that an arbitrarily low coefficient of variation cannot be attained, given $a_0 = 0$. However, as a general principle of insurance, the concept cannot be acceptable if the variability of premium payment is as random as the incidental claim itself. That would mean that the policyholder could not hedge his or her losses in exchange for a relatively flat price. Let $q_\lambda > 0$ be a value assigned to claim frequency constant λ, as a desired threshold of coefficient of variation. From the general formula it is evident that the coefficient of variation has its limits with respect to the boundedness:

$$s^2 = \frac{c}{2-c} \cdot \left(\frac{\underline{a}^\top \left(\sum_{k=1}^{2n+1} J_k \cdot B_{k-1}(\lambda)\right)\underline{a}}{\underline{a}^\top B_{(n)}\underline{\lambda}_{(n)}\underline{\lambda}_{(n)}^\top B_{(n)}^\top \underline{a}} - 1\right) < q_\lambda. \tag{3.22}$$

To solve the inequality for $\underline{a}$, it must be rearranged as

$$\underline{a}^\top \left(M_{(n)}(\lambda) - \left(1 + \frac{2-c}{c}q_\lambda\right) N_{(n)}(\lambda)\right)\underline{a} < 0, \tag{3.23}$$

where $M_{(n)}(\lambda)$ has already been defined and $N_{(n)}(\lambda) = B_{(n)}\underline{\lambda}_{(n)}\underline{\lambda}_{(n)}^\top B_{(n)}^\top$.

The matrix in the bracket is symmetric, however, it is not negative semidefinite and the existence of a solution cannot generally be expected. For instance, let $n = 1$ (or $f(x) = ax$), then $s^2 = \frac{c}{2-c}\frac{1}{\lambda}$, regardless of the value of a. If $c = 0.05$ and $\lambda = 10\%$, then the simplest system with $n = 1$ results in a coefficient of variation value 50.6%.

Similarly to corollary 3.14, in the case of Λ_t stationary Markov process, the low λ individual constant produces a high coefficient of variation.

In order to gain more insight into the actual coefficient of variation values, consider the following two models given the simplest claim frequency process $\Lambda_t \equiv \lambda$. Let the first one be an extremely trivial and intuitively useless one, where $r_{t+1} = r_t \times (1 - 0.9999) + d \times X_t$, which essentially means that the policyholder pays a fixed proportion of the previous period's claim amount. The second one is $r_{t+1} = r_t \times (1 - 0.05) + d \times (0.025 + X_t)$, in which there is a 5% decrease in the previous premium with a minor fixed increment and a proportional last claim amount. Observe the differences in magnitude shown in Table 3–1 for different values of λ. The first model (model 1) has $a_0 = 0$ in its $f(x)$ polynomial, implying that the coefficient of variation explodes for lower λ frequencies, see corollary 3.14. A low but positive $a_0 = 0.025$ constant was set in the second model (model 2), which results in a coefficient of variation tending to 0 as λ decreases around 0. The other important difference lies in parameter c, which forces the subsequent premium to override the previous one in the first model and exposes the payment entirely to the random fluctuation of claims, i.e. it is either approximately zero or d times the recent number of claims. In contrast, the second model does not allow the premium process to rapidly erase its past. This can only occur at a rate of $c = 5\%$, and consequently the relative variance of r_t compared to its average becomes substantially lower than in the first model. At the same time, the numerical example demonstrates the benefit of insurance, mitigating the policyholder's uncertainty with comparative values.

lambda	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
model 1	Inf	3.162	2.236	1.826	1.581	1.414	1.291	1.195	1.118	1.054	1
model 2	0	0.405	0.318	0.27	0.238	0.216	0.198	0.185	0.174	0.164	0.156

Table 3–1: The coefficient of variation in the two examples.

The coefficient of variation has been defined for the stationary premium process. However, taking into account the speed of convergence and the fact that stationarity might be reached on a longer run (depending on the value of c), the coefficient of variation as a function of time is a measure worth observing. Let this time-dependent coefficient of variation be $s_t = \frac{\sqrt{Var(r_t)}}{E(r_t)}$. It is easy to verify the following proposition.

Proposition 3.15 *The time-dependent coefficient of variation is*

$$s_t = \frac{d\sqrt{\sum_{k,l=1}^{t-1} (1-c)^{2(t-1)-k-l} \cdot \sigma_f(|k-l|)}}{r_1(1-c)^{t-1} + dE(f(X_1))\frac{1-(1-c)^{t-1}}{c}} \tag{3.24}$$

assuming that the state space of the stationary Markov chain λ_t is fixed. The numerator *is simplified for the* $\Lambda_t \equiv \lambda$ *case:*

$$s_t = \frac{d\sqrt{Var(f(X_1))}\sqrt{\frac{1-(1-c)^{2(t-1)}}{c(2-c)}}}{r_1(1-c)^{t-1} + dE(f(X_1))\frac{1-(1-c)^{t-1}}{c}} \tag{3.25}$$

and $\lim_{t\to\infty} s_t = s$ *in definition 3.4, given the usual assumptions for stationarity.*

3.2.5 Relative stationary average premium level

Definition 3.5 (relative stationary average premium level) Let

$$RSAL(\lambda) = \frac{\text{average premium level} - \text{minimum premium level}}{\text{maximum premium level} - \text{minimum premium level}}$$

denote the *relative stationary average premium level* as a function of the claim frequency.

In the present model, the premium can be arbitrarily close to zero or any positive constant, however, there is no theoretical maximum. The original definition is therefore not applicable unless some modification is introduced. Instead of taking the maximum premium level, let the denominator contain a number which is not exceeded with a *satisfying* probability. This is similar to the Norwegian example where there is no maximum premium, see [62].

Definition 3.6 (modified relative stationary average premium level) Let $RSAL^*(\lambda) = \frac{E(r^{*(\lambda)})-a_0d/c}{r_\lambda^\delta - a_0d/c}$ denote the *modified relative stationary average premium level* as a function of the claim frequency. Let $\delta > 0$ and r_λ^δ be defined as $\lim_{t\to\infty} P(\{r_t > r_\lambda^\delta\}) < \delta$, i.e. a level that is unlikely to be exceeded.

In other words, the definition is based on the stationary expected premium in the numerator, and the $1-\delta$-quantile of the stationary premium in the denominator. It is implicitly assumed that the minimum premium value is larger than zero, but can be arbitrarily close to zero if the constant coefficient a_0 in polynomial f is zero. If $a_0 > 0$, it is easy to prove that the minimum premium cannot go below value $\frac{a_0d}{c}$, and it may arbitrarily approach it as an asymptotical lower boundary. As an example, suppose that $\Lambda_t \equiv \lambda$, $f(x) = x$, $c = 0.1$, $d = 0.1$ and $\delta = 0.01\%$ in the definition. Table 3–2 presents $RSAL^*$ percentages for different policyholders with a claim frequency in the range of 10% to 100%. According to the RSAL table calculated for 30 countries in [62], given a claim frequency of 10%, the example model qualifies towards the top of the list. The higher this value, the better the policies are spread among different premium *classes*. The number of these classes are finite in most of the usual systems, however, the present model comprises an infinite amount of premium levels.

lambda	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
RSAL*	20.4%	27.5%	34.3%	36.4%	39.6%	44.3%	45.4%	46.9%	48.3%	50%

Table 3–2: Modified relative stationary average premium level as a function of claim frequencies.

3.2.6 Financial equilibrium

The stability of the total premium level in the entire population is crucial to the premium inflow's sustainability. The financial equilibrium of a system determined by transition rules over the years depends on the extent to which the average premium deviates from the original unit value. Such an equilibrium has been studied in [20]. If the proportion of more reliable policyholders, i.e. drivers with lower claim frequencies, is sufficiently large, then it may lead to a declining average premium level. A population with higher claim frequencies can conversely lead to steadily increasing premium levels. Due to the immature portfolio of new drivers, the financial situation can change rapidly over time, even if in the first years the total premium income of the insurance institutions is well-balanced. In that case, base premiums may need reassessment at some point in time, which presents the risk that a significant portion of the policies will lapse following the unfavourable change in price. It is therefore reasonable in the design of an experience rating scheme to ensure financial equilibrium, and to prevent a progressive decline or ascent in average prices.

Previously, we have studied the premium process r_t for an individual policy, i.e. $\Lambda_t = \lambda Y_t$ frequency process with a fixed $\lambda > 0$ constant and Y_t stationary Markov chain, where the stationary premium is $E(r^{*(\lambda)}) = E\left(r^* | \Lambda = \lambda\right)$. (Let Λ *without the subscript* t denote the random value of constant λ.) Suppose now that the claim frequency parameter λ is not a single value but can be from a set of K positive real numbers. This in turn reflects the segmentation of the policyholders into K groups, where in each group the drivers form a homogeneous risk community of similar capabilities. This in turn reflects the heterogeneity of risk in the entire population. The range of possible λ parameters should not necessarily be constrained to a finite set but rather may stem from an infinite selection or attain any positive real value. The distribution which governs λ is called mixing distribution, and can either be discrete or absolutely continuous, leading to several models for claim counts. If the λ parameter of process $\Lambda_t \equiv \lambda$ stems from Gamma(α, β), then the unconditional distribution becomes Negative Binomial $\left(\alpha, \frac{\beta}{1+\beta}\right)$, see [28]. Note that in regular bonus–malus systems, [57] delivers confidence intervals for premiums by bootstrapping and using nonparametric maximum likelihood estimation. Another example which has received attention in actuarial applications is the Delaporte distribution proposed by [25]. This is similarly a mixed Poisson with shifted gamma mixing distribution, see [88] for details of parameterisation.

Assume that the claim frequency parameter λ stems from a mixing distribution, and the objective is to find the unconditional distribution of the stationary premium level. Let Q stand for the mixing distribution, Υ for a random variable with distribution Q and $\underline{x}_{(n)}^\top = \begin{pmatrix} 1 & x & \dots & x^n \end{pmatrix}$. Furthermore, let

$$\underline{m}_{(n)}^\top = \begin{pmatrix} 1 & E(\Upsilon) & E(\Upsilon^2) & \dots & E(\Upsilon^n) \end{pmatrix}$$

contain the first n moments associated with distribution Q. The unconditional stationary premium level becomes

$$\begin{aligned} E(r^*) &= E\left(E(r^{*(\lambda)}|\Lambda)\right) \\ &= \int_{\mathbb{R}} \frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{x}_{(n)}) Z_{(n)} \pi' Q(dx) \\ &= \frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi', \end{aligned} \tag{3.26}$$

and simplifies to $E(r^*) = \frac{d}{c}\underline{a}^\top B_{(n)} \underline{m}_{(n)}$ if Λ_t is a constant process over time.

Observe the connection between the conditional and unconditional stationary premiums: in the first case, the matrix multiplication contains the diagonal matrix of the powers of λ, whilst in the second case, it includes the moments of the mixing distribution. If λ stems from distribution $Gamma(\alpha, \beta)$, then the vector of moments is $\underline{m}_{(n)}^\top = \begin{pmatrix} 1 & \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\cdot\beta} & \dots & \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\cdot\beta^n} \end{pmatrix}$. This explicit formula in the new model (to the extent of the gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$) facilitates the comparison of the stationary premium with the initial premium. Their equality means that there is no drift in the long term in terms of the aggregate expected premium volume of the entire in force population when compared to the initial r_1 price. This principle is defined in the theorem below.

Theorem 3.16 (Arató, Martinek) *If the initial premium level is*

$$r_1 = \frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi',$$

then the unconditional expectation of the premium remains r_1 *for any time in the* *future, i.e.* $E(E(r_t|\Lambda)) = r_1$ *for any* $t > 0$*. However, process* r_t *does not obey the martingale property.*

Proof. We shall prove by induction. By definition, we choose r_1 to be equal to $\frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi'$. Suppose that the expectation is true for t. Furthermore, we emphasise that r_t does not stand for the stationary price in the current case. Let $E(r_t) := E(E(r_t|\Lambda))$.

$$
\begin{aligned}
E(r_{t+1}) &= E(r_t)(1-c) + dE(f(X_t)) \\
&= \frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi' \cdot (1-c) + d \cdot \underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi' \\
&= \left(\frac{d}{c}(1-c) + d\right) \underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi' \\
&= \frac{d}{c}\underline{a}^\top B_{(n)} diag(\underline{m}_{(n)}) Z_{(n)} \pi',
\end{aligned}
\tag{3.27}
$$

hence, the expectations are equal for any $t > 0$.

For the second part of the statement, observe that $E(r_{t+1}|r_t) = r_t - cr_t + dE(f(X_t)) = r_t$ is true for $t = 1$, however, does not hold for $t > 1$. ■

Corollary 3.17 *As a special case of* $\Lambda_t \equiv \lambda$*,* $r_1 = \frac{d}{c}\underline{a}^\top B_{(n)} \underline{m}_{(n)}$ *satisfies theorem 3.16.*

For a special example, assume that $\Lambda_t \equiv \lambda$, $f(x) := a_1 x$ and the mixing distribution is $Gamma(\alpha, \beta)$, the unconditional stationary premium is then $E(r^*) = \frac{d \cdot a_1}{c} \frac{\alpha}{\beta}$. Parameters α and β are estimated on the basis of observations from the population, and c, d as well as the initial premium level r_1 can be selected arbitrarily. If the equality $r_1 = \frac{d \cdot a_1}{c} \frac{\alpha}{\beta}$ holds, the aforementioned stationary equilibrium of total premiums will hold as well. It must be emphasised that the unit value of the premium level can be arbitrary, as it reflects only a proportional value allocated to the policyholders. It is implicitly assumed that no policy lapses, so the distribution parameters characterising the underlying population are persistent accordingly.

3.2.7 Quadratic loss

Having discussed the concept of mixing distribution, the unconditional quadratic difference of the claim number and of the premium will now be investigated. For the sake of simplicity, we only elaborate on the $\Lambda \equiv \lambda$ case, i.e. when the stationary frequency process is constant *over time*. Let $h(t, \lambda) := E_\lambda (X_t - r_t)^2$ stand for the expected squared difference in period t. Furthermore, $h(\lambda) := \lim_{t \to \infty} h(t, \lambda)$ is a limit discrepancy that reflects the stationary premium and an independent claim amount, given a fixed λ claim frequency. Recall that $h(\lambda) = E_\lambda (X - r^*)^2$, where $X \sim \text{Poisson}(\lambda)$

and r^* stationary premium, are independent variables. Suppose that Λ is governed by the mixing distribution Q. For more condensed notations, let $u(\lambda) := E_\lambda f(X)$ and $v(\lambda) := E_\lambda f(X)^2$ be associated with the usual assumptions that $X \sim \text{Poisson}(\lambda)$ for a fixed λ claim frequency, and let $f(x)$ be the polynomial generating the autoregressive process.

Proposition 3.18 *Suppose that $\vartheta = \frac{d}{c}$ is the ratio of the two parameters in the recursive premium equation 3.3. The expectation of the squared difference for the average policyholder is*

$$E_Q h(\Lambda) = E_Q \Lambda^2 + E_Q \Lambda - 2E_Q(\Lambda u(\Lambda))\vartheta + \frac{2(1-c)E_Q u^2(\Lambda) + cE_Q v(\Lambda)}{2-c}\vartheta^2 \tag{3.28}$$

Proof.

$$\begin{aligned} h(\lambda) &= E_\lambda(X^2) - 2E_\lambda(X)E_\lambda(r^*) + E(r^{*^2}) \\ &= \lambda^2 + \lambda - 2\lambda \frac{d}{c} u(\lambda) + \frac{2(1-c)d^2u^2(\lambda) + d^2cv(\lambda)}{c^2(2-c)} \\ &= \lambda^2 + \lambda - 2\lambda u(\lambda)\vartheta + \frac{2(1-c)u^2(\lambda) + cv(\lambda)}{2-c}\vartheta^2 \end{aligned} \tag{3.29}$$

from proposition 3.4 and the proof of theorem 3.10. The expectation of $h(\Lambda)$ with respect to measure Q results in the proposition. ■

Corollary 3.19 *The parameter ratio $\vartheta = d/c$ minimising the quadratic loss function $E_Q h(\Lambda)$ satisfies*

$$\vartheta = \frac{E_Q(\Lambda u(\Lambda))}{\frac{2(1-c)E_Q u^2(\Lambda) + cE_Q v(\Lambda)}{2-c}}. \tag{3.30}$$

From the two equations 3.28 and 3.30 above it can be seen that the minimal quadratic loss is

$$\begin{aligned} E_Q h(\Lambda) &= E_Q \Lambda^2 + E_Q \Lambda - \frac{\left(E_Q(\Lambda u(\Lambda))\right)^2}{\frac{2(1-c)}{2-c} E_Q u^2(\Lambda) + \frac{c}{2-c} E_Q v(\Lambda)} \\ &= E_Q \Lambda^2 + E_Q \Lambda - \vartheta E_Q(\Lambda u(\Lambda)). \end{aligned} \tag{3.31}$$

Observe that the function is strictly monotonically increasing as a function of c, hence, its value tends to the minimum as c tends to the degenerate case zero. The reason for this is similar to that in theorem 3.10. Not only is the positivity of c required, but also from a practical perspective it must be reasonably large in order to ensure an appropriate pace of convergence to the stationary state. Therefore, it is proposed to first select the value of c and then to find the optimal parameters d and r_1 in line with the specified optimality criteria.

Let polynomial $f(x)$ of degree n and constant $c > 0$ be fixed as a speed of convergence parameter. The objective is to find d and initial premium r_1, creating a recursive model that minimises the expected squared difference $E_Q h(\Lambda)$ whilst satisfying the financial equilibrium constraint. In other words, find d and r_1 simultaneously satisfying equation 3.30 and $r_1 = \frac{d}{c}\underline{a}^\top B_{(n)}\underline{m}_{(n)}$ in corollary 3.17. This can be done in two separate steps. Let $\underline{m}_{(n)}^{+\top} := \begin{pmatrix} E(\Upsilon) & E(\Upsilon^2) & \dots & E(\Upsilon^{n+1}) \end{pmatrix}$ consist of the first $n+1$ moments of the mixing distribution Q. Furthermore, let $M_{(n)}^+$ be an $(n+1)\times(n+1)$ matrix, where the element (i,j) is the $(i+j-2)^{\text{th}}$ moment of Q. Matrix $M_{(n)}(\lambda)$ represents the one already shown in corollary 3.14. Reorganising equation 3.30 results in

$$d = \frac{c(2-c)\cdot \underline{a}^\top B_{(n)}\underline{m}_{(n)}^+}{2(1-c)\cdot \underline{a}^\top B_{(n)} M_{(n)}^+ B_{(n)}^\top \underline{a} + c\cdot \underline{a}^\top (E_Q M_{(n)}(\Lambda))\underline{a}}. \tag{3.32}$$

In the fraction $E_Q M_{(n)}(\Lambda)$ contains the element-wise expectation of the matrix with each element consisting of a Bell polynomial (Λ). With a minor calculation following corollary 3.14 it is concluded that

$$E_Q M_{(n)}(\Lambda) = \underline{a}^\top \left(\sum_{k=1}^{2n+1} J_k \cdot \left(\underline{e}_{2n+1,k}^\top B_{(2n)}\underline{m}_{(2n)} \right) \right) \underline{a}, \tag{3.33}$$

where $\underline{e}_{2n+1,k}$ is a column vector of length $2n+1$ with 1 in the kth element and 0 otherwise. The initial premium is as described in corollary 3.17, $r_1 = \frac{d}{c}\underline{a}^\top B_{(n)}\underline{m}_{(n)}$.

Now consider the case when the polynomial $f(x)$ and r_1 are fixed, and c, d are the unknown parameters. In this case c does not necessarily have a positive or even finite solution because parameter c has to satisfy

$$\frac{E_Q(\Lambda u(\Lambda))}{\frac{2(1-c)E_Q u^2(\Lambda) + cE_Q v(\Lambda)}{2-c}} = \frac{r_1}{\underline{a}^\top B_{(n)}\underline{m}_{(n)}}, \tag{3.34}$$

hence,

$$c = \frac{2\cdot\left(\left(\underline{a}^\top B_{(n)}\underline{m}_{(n)}^+\right)\cdot\left(\underline{a}^\top B_{(n)}\underline{m}_{(n)}\right) - r_1\cdot\left(\underline{a}^\top B_{(n)}M_{(n)}^+B_{(n)}^\top\underline{a}\right)\right)}{r_1\cdot\underline{a}^\top(E_Q M_{(n)}(\Lambda))\underline{a} + \left(\underline{a}^\top B_{(n)}\underline{m}_{(n)}^+\right)\cdot\left(\underline{a}^\top B_{(n)}\underline{m}_{(n)}\right) - 2r_1\cdot\left(\underline{a}^\top B_{(n)}M_{(n)}^+B_{(n)}^\top\underline{a}\right)},$$

where the denominator can be negative for some domain of r_1. This behaviour shows that r_1 cannot be arbitrarily selected if the lowest quadratic loss is required under the financial equilibrium constraint; for this reason it requires additional attention. The other unknown parameter d follows from c, r_1 according to the financial equilibrium equation, $d = \frac{c \cdot r_1}{\underline{a}^\top B_{(n)} \underline{m}_{(n)}}$.

Observe that different (c, d) and $(\tilde{c}, \tilde{d})$ pairs can result in identical quadratic loss and that modifying the parameterisation does not change the value of $E_Q h(\Lambda)$ subject to certain constraints. Suppose that $c, d, \tilde{c}$ are fixed and $\tilde{d}$ must be found in such a way that 3.28 remains invariant to the transformation $\tilde{c}, \tilde{d}$. From equation

$$\begin{aligned} -2E_Q(\Lambda u(\Lambda))\frac{d}{c} + \frac{2(1-c)E_Q u^2(\Lambda) + cE_Q v(\Lambda)}{2-c}\frac{d^2}{c^2} = \\ = -2E_Q(\Lambda u(\Lambda))\frac{\tilde{d}}{\tilde{c}} + \frac{2(1-\tilde{c})E_Q u^2(\Lambda) + \tilde{c}E_Q v(\Lambda)}{2-\tilde{c}}\frac{\tilde{d}^2}{\tilde{c}^2}, \end{aligned} \tag{3.35}$$

it can be seen that the associated quadratic equation (where $\tilde{d}$ is unknown) must have a non-negative discriminant in order to obtain a real solution. It can be shown that if the original pair c, d satisfies the minimising criterion 3.30 and $c < \tilde{c}$, then no real solution exists for $\tilde{d}$. Recall that the quadratic loss monotonically decreases as the value c decreases. Therefore, a similar loss value cannot be attained if the original pair c, d is optimal given c, and if $\tilde{c}$ is selected to be larger than c. To summarise this observation: if both parameter pairs $\{c, d\}$ and $\{\tilde{c}, \tilde{d}\}$ satisfy the local optimality 3.30, and $\tilde{c} > c$, then the quadratic loss associated with $\{\tilde{c}, \tilde{d}\}$ is strictly higher than the one associated with $\{c, d\}$. In contrast, if the parameters $\{c, d\}$ are not optimal in terms of equation 3.30, then alternative parameter pairs $\{\tilde{c}, \tilde{d}\}$ and $\{\tilde{\tilde{c}}, \tilde{\tilde{d}}\}$ exist, both of which satisfy equation 3.35 with the original c and d. Let the population be associated with Gamma(0.6,2.8) mixing distribution, for instance. Take $f(x) = 0.05 + x$, $c = 0.05$ and optimise d according to equation 3.30. Let $\tilde{c} = 0.04$, and find the value of $\tilde{d}$ in line with equation 3.35. Two solutions for $\tilde{d}$ are implied and hence, each of the premium processes

$$\begin{aligned} r_{t+1} &= 0.95r_t + 0.002 + 0.044X_t \\ \tilde{r}_{t+1} &= 0.96r_t + 0.002 + 0.038X_t \\ \tilde{\tilde{r}}_{t+1} &= 0.96r_t + 0.002 + 0.032X_t \end{aligned} \tag{3.36}$$

equally reflect an expected squared difference of $E_Q h(\Lambda) = 0.22$. To conclude the current section, the following proposition gives a generalisation of the quadratic loss function if Λ_t is an arbitrary stationary Markov chain.

Proposition 3.20 *If Λ_t frequency process is a stationary Markov chain, then the* *quadratic loss as a function of λ constant is*

$$
\begin{aligned}
h(\lambda) = & \underline{e}_{1,2}^{\top} diag(\lambda^0 \dots \lambda^n) Z_{(n)} \pi' - \\
& 2 \cdot d \cdot \sum_{k=1}^{\infty} \sum_{i=1}^{p} \sum_{j=1}^{p} \zeta_i \left(\underline{a}^{\top} B_{(n)} \underline{\zeta_j}_{(n)} \right) \pi_i (\mathbf{P}^k)_{i,j} \cdot (1-c)^{k-1} + \\
& \frac{d}{c(2-c)} \left(\sigma_f(0) + 2 \sum_{k=1}^{\infty} \sigma_f(k)(1-c)^k \right) + \left(\frac{d}{c} \underline{a}^{\top} B_{(n)} diag(\lambda^0 \dots \lambda^n) Z_{(n)} \pi' \right)^2
\end{aligned}
\tag{3.37}
$$

with the usual notations and

$$
\underline{e}_{1,2}^{\top} = \begin{pmatrix} 0 & 1 & 1 & 0 & \dots & 0 \end{pmatrix}, \quad \underline{\zeta_i}_{(n)} := \begin{pmatrix} 1 & \zeta_i & \zeta_i^2 & \dots & \zeta_i^n \end{pmatrix}^{\top}.
$$

Proof. The decomposition of $h(\lambda)$ is

$$
E\left(X_t - r_t^{*(\lambda)} \right)^2 = EX_t^2 - 2E(X_t r_t^{*(\lambda)}) + E(r_t^{*(\lambda)})^2.
$$

The first and the second term follows from theorem 3.5 and 3.10, the only unknown term is the middle one. Similarly to the proof of proposition 3.12,

$$
\begin{aligned}
E(X_t r_t^{*(\lambda)}) &= E\left(X_t \cdot \sum_{k=1}^{\infty} f(X_{t+k}) d(1-c)^{k-1} \right) \\
&= d \sum_{k=1}^{\infty} E(X_t f(X_{t+k}))(1-c)^{k-1} \\
&= d \sum_{k=1}^{\infty} \sum_{i=1}^{p} \sum_{j=1}^{p} \zeta_i \left(\underline{a}^{\top} B_{(n)} \underline{\zeta_j}_{(n)} \right) \pi_i (\mathbf{P}^k)_{i,j} \cdot (1-c)^{k-1},
\end{aligned}
\tag{3.38}
$$

which proves the proposition. ■

3.3 Comparison with existing schemes

3.3.1 Schemes applied in Europe

For comparative purposes and in order to highlight the practical relevance, the proposed model was explored in the context of existing schemes that have been in place in Europe for decades. The present chapter does not aim to describe the various bonus–malus systems in-depth, as this has already been done in a wide range of literature. The models selected for comparison were from the Belgian, the Hungarian and the Dutch systems as most of them have been extensively studied in the actuarial literature, see [60].

Recall that these European schemes consist of a finite number of premium classes, and transition rules are defined as a function of two elements: (1) claims generated in a period and (2) previous class. In other words, the premium process is a Markov chain (assumption 2.4) and a transition probability matrix determines the probabilities of the classes in a subsequent period, which in turn are dependent on the class in the previous period and the claim frequency parameter. Each insured person is initially placed into a deterministic class for the first year.

With respect to the Belgian model, this comparison follows the transition rules and relativities as described in [62]. In the case of the Hungarian model, [24] specifies the rules. Finally, for comparison with the Dutch system, the most recent rules obtained from the website of a Dutch insurance institution [37] are applied; these are not identical to the versions referred to in older papers. The relativities are shown below and are listed in order of states with the highest to lowest surcharge within each country. Guiding regulations in all three of the countries allow insurance institutions to determine their own premium levels in each class but stipulate that the transition rules and classes remain unchanged.[1]

- Belgium: 2.00, 1.60, 1.40, 1.30, 1.23, 1.17, 1.11, 1.05, 1.00, 0.95, 0.90, 0.85, 0.81, 0.77, 0.73, 0.69, 0.66, 0.63, 0.60, 0.57, 0.54, 0.54, 0.54, with an initial premium level of 0.85 (classes: $22, \dots, 0$, initial class: 11). Transition rules: claim-free year -1 class (special rule: after 4 consecutive claim-free years the premium level cannot be above 100), first claim +4 classes and further claims +5 classes in one year. In spite of the special rule, [86] has shown that the system represents a Markov chain.

- Hungary: 2.00, 1.60, 1.35, 1.15, 1.00, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50, with initial premium level 1 (classes: $15, \dots, 1$, initial class: 11). Transition rules: claim-free year -1 class, each claim +2 classes, more than 3 claims in a year results in moving to the highest malus class.

- The Netherlands: 1.15, 0.95, 0.85, 0.75, 0.65, 0.60, 0.55, 0.50, 0.45, 0.40, 0.35, 0.30, 0.28, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, with initial premium level 0.6 (classes: $22, \dots, 0[, -1, \dots]$, initial class: 17). Transition rules: claim-free year -1, each claim +5 classes. (The long sequence of low premium levels at the end represents the consolidation of discount after several claim-free years.)

[1] The Commission of the European Communities filed a lawsuit against the French Republic in 2002 claiming that the enforcement of the bonus–malus system was limiting free competition (case C-347/02). The court ruled in favour of France in 2004.

3.3.2 Comparative overview

All calculations have been programmed in R. A set of claim histories for 1,000 policyholders is generated over 30 years. This insured population is put into both the newly proposed model as well as the already existing bonus–malus systems for comparison purposes.

Real-life data published by the Macquarie University have been used for the parameterisation of the mixing distribution, see also [48]. The source contains 67,856 vehicle policies from 2004-2005, each covering a period of one year. There are 4,624 policyholders who reported at least one claim, with a maximum of 4 claims. If it is supposed that the simplest frequency process which is constant over time, i.e. the Poisson-Gamma model is assumed to be valid, then it implicitly means that the unconditional claim numbers are negative binomial variables. We construct a version of the method of moments and solve the following system of equations (see definition 2.3 with $t_i = 1\ \forall i$):

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{\sum_{i=1}^{m} Z_i}{m} \quad \text{and} \quad \frac{1}{\hat{\beta}} = \left(\frac{\sum_{i=1}^{m} Z_i^2}{\sum_{i=1}^{m} Z_i} - 1 \right) - \frac{\sum_{i=1}^{m} Z_i}{m}, \tag{3.39}$$

Z_i $(i = 1, \ldots, m)$ are the individual claim numbers and m is the total number of policies. Solving the equations results in $\hat{\alpha} = 1.141$ and $\hat{\beta} = 15.683$.

However, the method requires further discussion in the general stationary Markov chain frequency process. Assume that the distribution of the individual Λ parameter is Gamma. Let Y_t be a random walk on three ζ_i states: 0.2229, 1.1145, 2.2289, with probabilities of stepping left and right $\mu = 0.4$ and $\nu = 0.3$, see figure 3–1. It can be calculated from the known formulas that $EY_t = 1$.

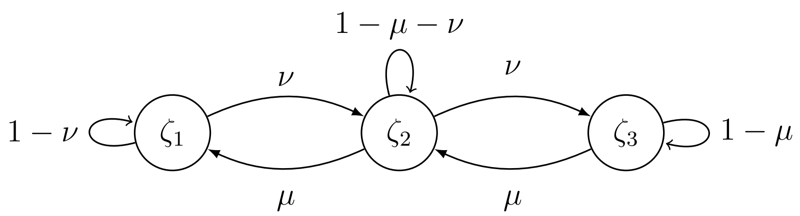


Figure 3–1: The graph of random walk Y_t.

The discussed properties of X_t can be used to calculate

$$E(X_t) = E(E(X_t|\Lambda_t)) = E(\Lambda_t) = E(E(\Lambda_t|\Lambda)) = E(\Lambda E(Y_t)) = E(\Lambda) = \frac{\alpha}{\beta} \tag{3.40}$$

and

$$\begin{aligned} E(X_t^2) &= E(E(X_t^2|\Lambda_t)) = E(\Lambda_t^2 + \Lambda_t) = \\ &= E(\Lambda^2) \cdot \sum_{k=1}^{p} \pi_i \zeta_i^2 + E(\Lambda) = \frac{\alpha}{\beta} + \sum_{k=1}^{p} \pi_i \zeta_i^2 \frac{\alpha(\alpha+1)}{\beta^2}, \end{aligned} \tag{3.41}$$

which defines a system of equations associated with a method of moments parameter estimation. Let y_2 stand for $\sum\limits_{k=1}^{p} \pi_i \zeta_i^2$, then

$$\hat{\beta} = \frac{\frac{\sum_{i=1}^{m} Z_i y_2}{m}}{\frac{\sum_{i=1}^{m} Z_i^2}{m} - \frac{\sum_{i=1}^{m} Z_i}{m} - \left(\frac{\sum_{i=1}^{m} Z_i}{m}\right)^2 y_2^2} \quad \text{and} \quad \hat{\alpha} = \frac{\sum_{i=1}^{m} Z_i}{m} \hat{\beta}. \tag{3.42}$$

The first moment $\frac{\alpha}{\beta}$ of Λ does not change compared to the case which is constant over time, however, the variance increases as the Λ_t process becomes non-constant over time. The new parameters in the proposed example are $\hat{\alpha}' = 0.241$ and $\hat{\beta}' = 3.306$. The settings of process Y_t require thorough scrutiny of the underlying phenomena affecting the insureds' claim reports, for instance, the impact of weather conditions on the drivers' performance year by year. The remainder of this section assumes that $Y_t \equiv 1$.

The extent to which the analysed schemes are financially balanced is subject to examination. Figure 3–2 represents an insured population of 1,000 policies over 30 years. Suppose that the claim frequencies are randomly and independently selected from distribution $Gamma(1.141, 15.683)$, i.e. the mean claim frequency is 0.073. In the recursive model, parameters are chosen as follows: the initial premium is $r_1 = 1$, $c = 0.05$, $d = 0.687$ and $f(x) = x$, or otherwise, the recursion follows $r_{t+1} = 0.95 r_t + 0.687 X_t$. The Belgian, the Hungarian and the Dutch systems reflect a higher aggregate premium level at the launch of the portfolio, which significantly decreases during the first 10-15 years. There is no significant reason for arbitrarily choosing $f(x) = x$ to illustrate the total premium level over the years. The goal of the example is to show the stability of premium in-flows. If triplet $\{c, d, r_1\}$ satisfies the equality in theorem 3.16, then the expected value of the premium remains constant over time, regardless of the driving $f(x)$ polynomial. Similar results could be plotted for $f(x) = 4x + 3x^2 + 5x^3$.

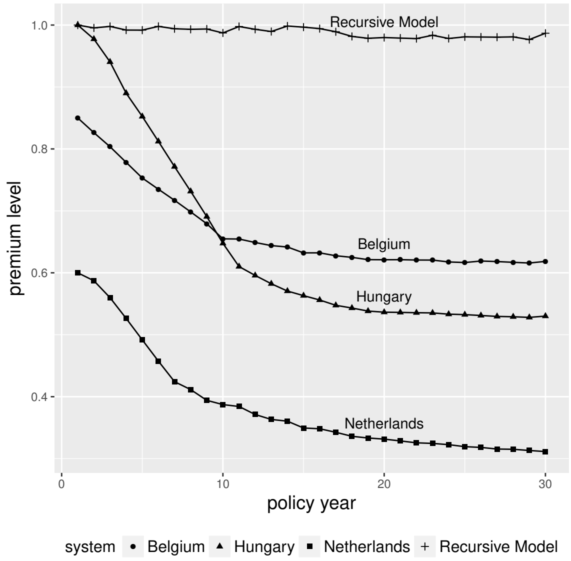


Figure 3–2: Total premium level of a portfolio in the first 30 policy years.

There is an important difference between the preliminary settings of the new recursive model and the existing three models. In the proposed model the parameters are set on the basis of population constraints, thereby embedding the premiss that the claim frequency is governed by the Gamma distribution with parameters α and β. In contrast, the existing bonus–malus schemes do not use this information, which makes the comparison biased. In order to obtain a more appropriate comparison, three basic features of the existing systems can be considered as changing attributes: (1) transition rules, (2) relativities and (3) initial relativity class. (1) Any amendment with respect to transition rules results in a substantial shift in the core properties of a scheme, as modifying the transition rules creates a new system with a new stationary distribution. (2) Defining new relativities or premium levels does not affect the limit distribution, however, it affects stationary premium and the metrics presented in the previous sections. (3) We propose to change the initial relativity class so that the initial premium level is as close as possible to the stationary expected one. The premium is asymptotically insensitive to the initial level.

In an ordinary bonus–malus scheme, let $M_O(\lambda)$ denote the squared $(m \times m)$ matrix of transition probabilities as a function of claim frequency, where m stands for the number of separate classes. The stationary distribution $\underline{\pi}_\lambda$ is the eigenvector of $M_O(\lambda)$ corresponding to the eigenvalue 1. Each of the elements represents a premium class in the system. If the array of premiums is $\underline{u} = \left(u_1, \ldots, u_m\right)^\top$, the expected stationary premium u^* given claim frequency λ is $E(u^*|\lambda) = \underline{\pi}_\lambda^\top \underline{u}$. Furthermore, the unconditional expectation can be expressed as

$$E(u^*) = E(E(u^*|\Lambda)) \approx \frac{1}{N}\sum_{i=1}^{N} \underline{\pi}_{\lambda_i}^\top \underline{u} = \left(\frac{1}{N}\sum_{i=1}^{N} \underline{\pi}_{\lambda_i}^\top\right) \underline{u}, \tag{3.43}$$

where $\lambda_1, \ldots, \lambda_N$ is an i.i.d. sample from the mixing distribution (Gamma(α, β) in the examples). Hence, if the initial state in the system is as close as possible to $E(u^*)$ without changing the original u_i prices, the initial price will be approximately equal to the convergence point of the expected premium. In other words, instead of the original initial premium class *init*, let the first class be *init*$'$, where $u_{init'} \approx E(u^*)$. This step is based on the assumption that the population has already been observed and that the claim frequency distribution has been evaluated in preliminary research.

As an example, figure 3–3 presents a comparison of average premium trajectories for a portfolio of 10,000 policyholders. Since each model starts at a different relativity level, the curves are normalised so that they each start at 100% and can be compared. The claim frequency distribution is assumed to be governed by the Gamma distribution with $\alpha = 1.141, \beta = 15.683$. Let the parameters of the recursive model be $c = 0.04, d = 0.03$, $f(x) = 0.025 + x$, i.e. the recursion is

$$r_{t+1} = 0.96 r_t + 0.001 + 0.03 X_t. \tag{3.44}$$

Parameter d and initial premium r_1 are optimised as described in theorem 3.16 and section 3.2.7. All of the observed schemes are designed to converge to the initial state on average. The extent to which the models can show financially balanced characteristics can be observed from the average premium trajectories in the years between the starting point and the endpoint. It is important to emphasise that the path does not (or to a very small extent, depending on the number of policies) reflect a randomness of premiums, rather it illustrates the way the premium of the population changes, which is *embedded into the transition rules and premium levels* of the original schemes. In other words, it shows how intensely the average relativity changes over time before approaching the convergence level given the claim frequency distribution.

The Belgian scheme has an expected stationary premium of 0.601, corresponding to the initial class no. 4, which reflects an initial premium level of 0.6. Similarly, in the case of Hungary, the expected stationary premium is 0.526. Let the initial class be no. 2 and the starting premium level of the class be 0.55. For the Netherlands, the expected stationary premium is 0.293. Let the initial class be no. 11, and the starting

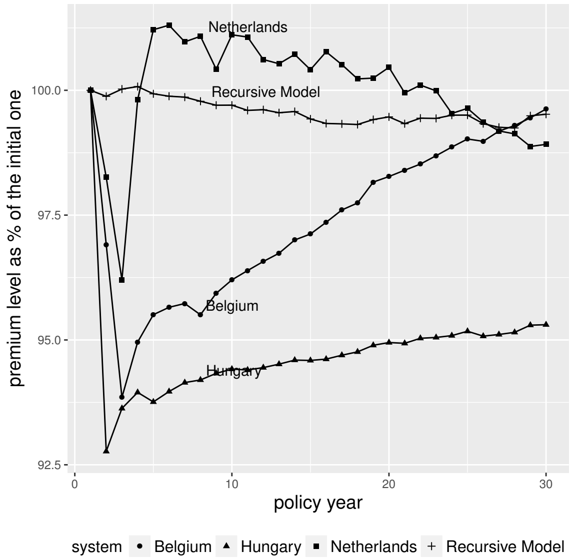


Figure 3–3: Total premium level of a portfolio in the first 30 policy years with initial premium levels, adjusted in accordance with the population claim frequency distribution (% of the initial premium).

premium level of the class be 0.3. The squared difference as a function of λ can be found in table 3–3, which is calculated on the basis of corollary 3.7 for the recursive model. This table is not suitable for the comparison of different models and must be evaluated separately due to the fact that relativities are determined differently for each country and for the recursive model. A lower squared difference on average can still represent a model with better predictive qualities.

	Recursive	Belgium	Hungary	Netherlands
10%	4e-05	0.23497	0.17871	0.02734
20%	0.00098	0.42179	0.14625	0.10243
30%	0.00318	0.82612	0.22288	0.21299
40%	0.00664	1.06359	0.49647	0.22347
50%	0.01135	1.12976	0.76048	0.18822
60%	0.01732	1.099	0.87042	0.13957
70%	0.02454	1.01633	0.87081	0.09127
80%	0.03302	0.90721	0.81231	0.05012
90%	0.04275	0.78653	0.72389	0.01992
100%	0.05374	0.66338	0.62187	0.00305

Table 3–3: Squared difference between the stationary premium level and the expected claim frequency as a function of λ multiplier of the stationary process (corollary 3.7).

Elasticities are compared in Figure 3–4. Observe that the recursive model (Rec) is not associated with a constant 1 elasticity function due to the non-zero a_0 term in the recursion polynomial. Fixing a minimum premium requirement of $\frac{a_0 d}{c} = 0.019$ places a relatively strict penalty on an insured person with an extremely low number of expected claims, an attribute inherent to any other model. As a direct corollary of proposition 3.8, the function tends to 0 as $\lambda \to 0$. With respect to this metric, the new scheme performs better than the existing ones for the majority of claim frequencies, i.e. its elasticity tends to 1 strictly monotonically (as $\lambda \to \infty$) and increases rapidly close to zero.

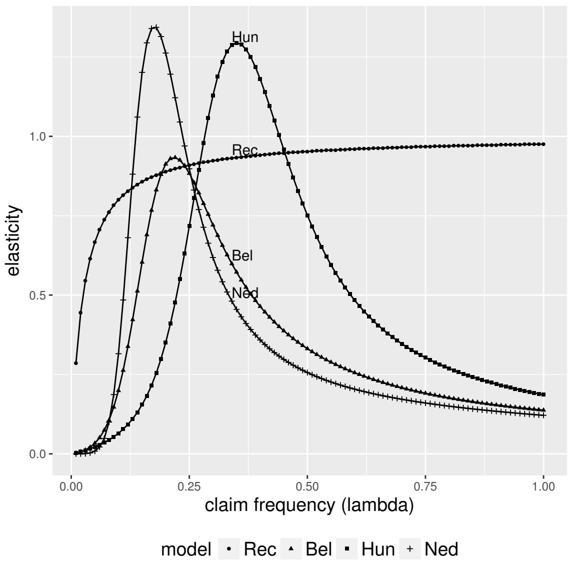


Figure 3–4: Elasticity as a function of claim frequency.

The coefficients of variation are compared in Figure 3–5. Corollary 3.14 and $a_0 > 0$ ensure that this value tends to zero as the claim frequency declines in the recursive model. For the range of policyholders with better claim frequencies, i.e. for those with a λ lower than approximately 0.1, the coefficient of variation is higher in the new system than in the referenced ones. Thus, from the perspective of relative volatility in price, the autoregressive model performs more poorly. On the contrary, for claim frequency parameters above approximately 0.1, the coefficient intersects the Belgian and Hungarian curves from above and subsequently remains below them. For the range of policyholders representing a fairly risky group, for $\lambda > 0.3$, the recursive model's coefficient of variation is closest to that of the Dutch system.

Figure 3–6 shows the evolution of the coefficient of variation over a period of 30 years for a policyholder with claim frequency $\lambda = 0.1$. The speed of convergence needs attention, as all the observed models are accompanied with some volatility in this measure, even after decades. In the first 4 years, the new recursive model has a lower coefficient than its Dutch and Hungarian competitors. This has a high degree of practical relevance if the average policyholder spends less than 4-5 years in the system. After 4 years, the curve of the new model continues to increase monotonically, converging to its stationary value of 0.361 (also see figure 3–5 for $\lambda = 0.1$). Schemes of

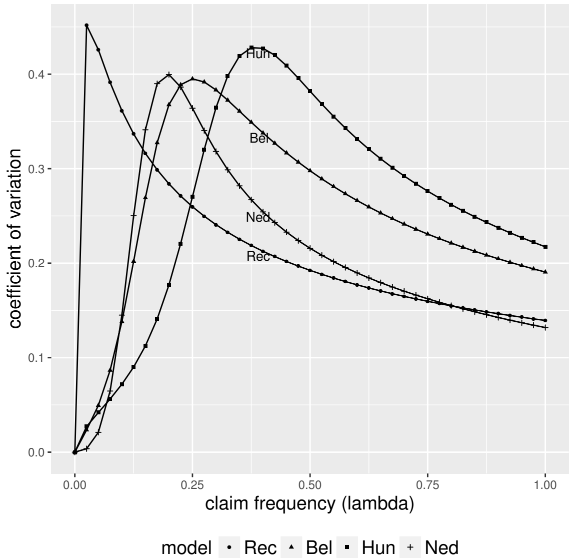


Figure 3–5: Coefficient of variation as a function of claim frequency.

several other countries result in even higher figures, e.g. the Swiss one is approximately 0.5 in the time range of 10-30 years, see [62]. Note that the observed population consists of a relatively low risk community with an average claim number of 0.073. Suppose that this average is higher and take 0.21, for instance. As a consequence of a higher average λ value, the coefficient of variation becomes the lowest during the first 10 years in the recursive model compared to the original models, given $\lambda = 0.1$. Thus, on similar curves as in figure 3–6, the recursive one runs below the other ones in the first decade. Two reasons explain this difference between the curves associated with the averages of 0.073 and 0.21. The first one is that in the original models, we have chosen the initial state such that the financial equilibrium constraint is satisfied, hence, with a modified initial class the coefficient of variation changes as well. Indeed, the coefficient is altered in the first years due to the random walk starting in a different initial state. The second reason is that the parameterisation of the autoregressive model depends on the mixing distribution parameters α and β, as c, d, r_1 is a result of the optimisation with respect to the financial equilibrium constraint as well as the quadratic loss. These two reasons explain why the average λ of the population affects the coefficient of variation as a function of time, even for a fixed claim frequency of 0.1.

Table 3–4 contains the relative stationary average premium levels (RSAL) given claim frequency parameters between $10\% - 100\%$. Systems of RSAL exceeding 20% are rare; some of the few examples include Kenya, Spain and Malaysia, see [61]. Higher values can be interpreted as better distributed premium levels among policyholders.

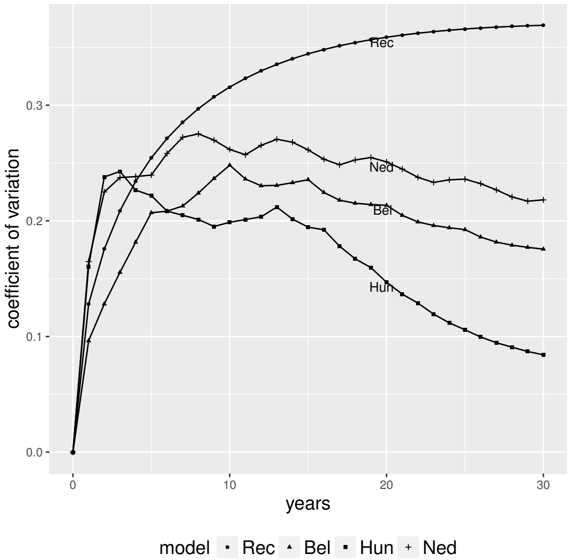


Figure 3–6: Coefficient of variation as a function of time ($\lambda = 0.1$).

In contrast, lower values reflect clusters of the insured in more benign bonus ranges. Note that for the recursive model the modified definition 3.6 has been applied with $\delta = 0.01\%$, due to the lack of a theoretical maximum relativity. For the three countries and for a $\lambda = 10\%$ claim frequency the RSAL is fairly low and for Hungary it remains below 20% even for a relatively risky constant $\lambda = 30\%$. RSAL detects that policies associated with such λ values experience a high clustering in the better bonus classes. This clustering at the stationary state implies that an excessive penalty is applied for new policyholders and that the average premium amount inevitably decreases as the steady state is approached. The recursive model shows a wider spread of policies among the premium levels, as is evident for $\lambda = 10\%$ where the RSAL is above 30%. Since the claim frequency range between 0% and 30% covers the majority of drivers in automobile insurance, the comparison of RSAL values through these frequencies is emphasised.

3.4 Conclusion

A new experience rating model for non-life insurance has been constructed that is particularly applicable to motor third-party liability insurance. In contrast to the currently used bonus–malus systems based on random walks on the state space of relativities, this new approach identifies a first-order autoregressive model. In this process the role of white noise is played by a function of the number of individual claims that occurred in a given period. We have proved the existence of the stationary distribution under sufficiently general circumstances, and drafted partially explicit

	Recursive	Belgium	Hungary	Netherlands
10%	31.53%	3.06%	1.52%	1.71%
20%	42.27%	21.2%	5.49%	30.01%
30%	46.08%	45.82%	18.14%	56.83%
40%	51.79%	61.05%	40.31%	69.19%
50%	55.55%	70.06%	58.14%	75.98%
60%	57.46%	75.91%	68.86%	80.4%
70%	59.59%	80.01%	75.54%	83.57%
80%	60.81%	83.05%	80.09%	85.98%
90%	61.65%	85.4%	83.39%	87.9%
100%	63.97%	87.29%	85.91%	89.47%

Table 3–4: Comparison of relative stationary average premium levels as a function of claim frequencies ($\lambda = 10\% - 100\%$).

formulas for elasticity, coefficient of variation, modified relative stationary average premium level and financial equilibrium. The concept has been constructed under the assumption that the X_t claim numbers are conditionally independent Poisson random variables given the λ_t frequency parameters, where the frequency process is governed by a stationary Markov chain.

One section in this chapter is dedicated to a comparative analysis between the new model and three existing European schemes. With the new model, financial equilibrium and elasticity can arbitrarily be set to optimal, thereby outperforming the other models. In terms of the coefficient of variation, the autoregressive model might underperform the others in the range of low claim frequencies, yet may perform better for slightly riskier policies with claim frequency parameters above 0.1-0.2. RSAL is consistently good, especially for policyholders with a low number of expected claims.

Average optimal retention, frequently referred to as the hunger for bonus phenomenon can be tested in further research, see [61], [18]. Another way of extending the present results is to investigate the model given other distributions governing claim counts, see [112]. Lastly, insurance products different from automobile liability can be considered in the framework of the new model, with transitions driven by the total claims paid rather than by claim numbers. This may lead to a more effective exploitation of larger degree polynomials in the recursive formula.

Part II

Comparison and Ranking of Models in Stochastic Claims Reserving

Glendower *I can call spirits from the vasty deep.*
Hotspur *Why, so can I, or so can any man;*
But will they come when you do call for them?

— **William Shakespeare**
Henry IV, Part I, Act 3, Scene 1

4 Simulation-Based Comparison of Stochastic Claims Reserving Models

4.1 Introduction to stochastic reserving

Insurance and reinsurance institutions, particularly property and casualty insurers, put a considerable amount of effort into understanding outstanding claims reserves. These amount to the most material proportion of technical reserves[1], hence, it is critical that actuaries and management control their volume and uncertainty. It is not only the measure and pattern of future cash outflows and metrics of associated risks that play a role in the insurance business; management decisions are also triggered by the outcome of calculations.

The appropriate estimation of incurred but not reported (IBNR) and reported but not settled (RBNS) outstanding claims is crucial to preserve the solvency of insurance institutions, especially in the property and casualty business. A general assumption is that claims related to policyholders are reported to insurance companies in subsequent years, sometimes several years later. Often, the payout is delayed as well. Thus, reserves have to be made to cover the arising obligations.

In principle, stochastic claims reserving has been developed in order to capture the distribution of claims incurred. Hence, in contrast to traditional point or deterministic estimations, in a stochastic manner more information can be explored with respect to the nature of future occurrences. Whilst a simple point estimation is not suitable to address the tail or more extreme events, stochastic reserving can predict the claim payments related to rare events beyond the average estimation. This allows actuaries to analyse the variance of future payments. For extensive descriptions of stochastic reserving see [35, 111], and for description of development models and evaluation techniques see [102, 69, 107]. These methods not only estimate the expected

[1] The non-life insurance technical reserves of the euro area insurance corporations at the year-end of 2017 was EUR 523 bln [7].

value of the outstanding claims, but also examine their stochastic behaviour. Most of the stochastic reserving methods estimate the distribution of outstanding claims, hence, they can be considered as probabilistic forecasts.

Scholars and industry professionals have been studying different estimation models for the past decades extensively. By now, interest in stochastic models has outgrown the interest in deterministic ones, shifting from simple point estimations to an approximation of probability distributions: This means that features of the examined object can be calculated with more insight into the nature of the underlying phenomenon. The demand for forecasts embodied in distributional forms rather than point estimates has grown rapidly along with the growth of computational power, simultaneously allowing for the pragmatic implementation of Monte Carlo type algorithms. This increasing interest has emerged not only in insurance but in several other disciplines, such as meteorology or finance, demanding a more meaningful prediction of future outcomes. [35, 111] contain comprehensive overviews of reserving methods. In our view, the validation of the models on actual industrial data and the comparison of these models' appropriateness is a crucial question. Next to the relevance of model suitability, proportionally to the size of existing literature on models, even more attention has to be given to the substantiation of model quality and to the comparison of methods. Professionals who are offered countless different models need guidelines that can support an optimal selection. A more recent work, [77] performs investigation on bootstrap and Bayesian models using publicly available claims data from American insurance companies. The work also proposes new methods practically solved through MCMC simulations.

A case study is performed in [109] in order to analyse accounting year effects in run-off triangles.[2] This study compares Bayesian models with mean square error of prediction (MSEP) and deviance information criterion (DIC). [96] and [98] provide another alternative with Q-Q plots and P-P plots. Nevertheless, the first one focusses on understanding the dependency among the triangles of different business lines with a copula regression model, and the second one describes retrospective tests on the proposed models. Even more focus is put on the validation of methods in [74], evaluating which method should be preferred. Three methods, the double chain ladder, the Bornhuetter–Ferguson and the incurred double chain ladder methods are compared through two real data sets from property and casualty insurers, and the metrics used are call error, calendar year error and total error. Supported by real-life claims data, [104] compares three models with different residual adjustments using

[2] A ubiquitous concept in Part II that represents the claim observations in a triangle-shaped form.

the Dawid-Sebastiani scoring rule (DSS).

This chapter analyses diverse stochastic claims reserving methods by means of several goodness-of-fit measures proposed by the authors of [4]. In a game-theoretic interpretation of forecasts, it sets up a ranking framework that selects from competing models. Certainly, there is hardly any ranking technique which all actuaries would unanimously accept, as a peremptory selector of the most proper prediction models. However, it is reasonable to define and observe the important characteristics of estimations, which put together may support the decision-making process and the validation of applied methods. In the assessment of reserving models, there is a strong intention to promote measures originally used in stochastic forecasting. Probability integral transform provides more justification on the predictive distribution appropriateness, whereas the Kolmogorov–Smirnov or Cramér–von Mises statistics would fail to shed light on what exactly goes wrong with the hypothesis. Established scores compare and verify qualities of rival probabilistic forecasting models on the basis of estimation and real outcomes.

From the wide range of scoring rules we apply the continuous ranked probability score and its generalisation, the energy score, due to their flexible applicability on differing distributions, see [46]. Coverage shows the central prediction interval of a prediction given a real governing distribution. Sharpness stands for the width as expected difference between lower and upper p-quantiles, the narrower the better expressed in payment, see [47].

Concerning several stochastic estimation methods, the comparison of their appropriateness and distinction making approach is introduced, supposing different distributional development models. The chapter presents simulations on the basis of various run-off triangles gathered from actuarial literature. Calculations have also been run on a real world data set from an insurance company, generating 2000 scenarios with random draw from the claims, and determining the most appropriate stochastic reserving methods. Only paid run-off triangles have been used, thus making comparisons with the MCMC model described in [76] was technically feasible. Note that some methods are out of scope of the present chapter, such as bootstrap methods in [11] and [85], and the MCMC method from [71].

Due to the lack of explicit expression for the distribution of the ultimate claim values, the Monte Carlo type evaluation is inevitable in a considerable amount of cases. For instance, the convolution of random variables governed by log-normal distribution cannot be handled in an analytical form. The proposed comparison method has been applied on an itemised (claims and payouts) dataset from an insurance company, where both the upper and the lower triangles were known. In a multistage simulation-based framework, 2000 scenarios have been created with random draw and replacement from the claims, followed by the selection of the most appropriate stochastic reserving methods based on these scenarios.

Section 4.2 contains an introduction to five frequently applied parametric models. The ratio of cumulative claims stemming from subsequent years might follow log-normal distribution, which has an empirical background from a large amount of data analysed in practice. In addition, negative binomial, Poisson, overdispersed Poisson and gamma distributed models are addressed. Section 4.3 describes data used for calculations, and how scenarios can be defined for model fitting in the stochastic reserving framework. In section 4.4, different stochastic claims reserving techniques are explained. In addition to the standard comparison measures - i.e. what is accepted in industry practice - section 4.5 widens the set of such comparison tools with the continuous ranked probability scores, along with an introductory example. Section 4.6 establishes a simulation-based algorithm and explains how to interpret results. Finally, section 4.7 concludes the chapter. The chapter is based on [2, 4, 72].

4.2 Models with underlying distribution

Data (including future claims) are usually represented by $I \times J$ matrices, where element $X_{i,j}$ $(i, j = 1, 2, \ldots)$ represents the claim amount incurred in year i, and paid with a delay of $j - 1$ years.

Definition 4.1 (run-off triangle − incremental) Data consisting of the past observations is the part of the matrix including and above the anti-diagonal, $X_{i,j}$ $i+j \leq I + 1$, often referred to as *run-off triangles* or *upper triangles*.

In the context of Chapter 4 and 5, matrix diagonal will always mean the anti-diagonal. Run-off triangles incorporate the data of claims already incurred and administered by the insurance institution, and the upper triangle synonym stems from the fact that the run-off triangle is the subset of the claim matrix above (and including) the anti-diagonal. Taking one row into consideration, each subsequent element means that $X_{i,j+1}$ additional claims incurred in year i, increasing the already accumulated amount of $X_{i,1} + \ldots + X_{i,j}$. Hence, this run-off triangle representation is called with its elements incremental and $X_{i,j}$ are the incremental values.

Definition 4.2 (run-off triangle − cumulative) We may also define a triangle with the aggregate claim values $C_{i,j} := X_{i,1} + \ldots + X_{i,j}$. Without loss of generality, assume that $I = J$.

Unknown elements for indices $i + j > I + 1$ have to be predicted, which is the lower triangle part of the matrix defined above and which represents future claims. For the sake of precision, two distinct data types can build up a loss triangle: one is the *incurred claims* and the other one is the *paid claims*. Incurred claims refer to amounts which (1) have been reported to the insurer and booked in the administration system as outstanding liabilities (case reserves or RBNS) and which (2) have not yet been reported (IBNR). These amounts may change some time after they have been booked due to the reassessment of claims. Paid claims represent the actual payments and they are usually not modified afterwards. The total estimated liability the company holds is the aggregate incurred claims decreased by the actual paid amount

$$\text{Reserve} = \sum_{i=1}^{I} C_{i,I}^{\text{Inc}} - \sum_{i=1}^{I} C_{i,I-i+1}^{\text{Paid}}.$$

Definition 4.3 (ultimate claim) Define *ultimate claim* value as the sum of observed (upper triangle) and outstanding claims (lower triangle). It can either relate to the incurred claims $UC^{\text{Inc}} = \sum_{i=1}^{I} C_{i,I}^{\text{Inc}}$ or to the paid claims $UC^{\text{Paid}} = \sum_{i=1}^{I} C_{i,I}^{\text{Paid}}$. (Superscripts Inc and Paid are omitted unless they are necessary.) Assume that there is no claim payment beyond development year I, hence, $X_{i,j} = 0$ if $j > I$.

The difficulty of claims reserving is the prediction of UC, row-wise $C_{i,I}$ or even all elements $X_{i,j},\ i + j > I + 1$. Results in this chapter are calculated on the total ultimate claim value UC. However, predictions might be shown separately for each occurrence year (applied on the rows). The generality of the chapter's results are not violated if the reserves are segmented into occurrence years rather than looked at on an aggregate level.

Five parametric development models are characterised in the following section, applied in the insurance industry. Depending on the nature of the set of insurance contracts, different assumptions are empirically justifiable. This means that reporting times related to fire insurance should be handled separately from liability insurance with typically long run-offs. Whilst the former claims are reported and settled usually within 2 years, liability related claims have a long latent period.[3] In other words, the distribution and indirectly the evolution of $X_{i,j}$ (or $C_{i,j}$, in aggregate) values are supposed to be very diverse for various contract types. In this section, five ways of looking at the evolution of reported claims are discussed. These are conventional in the sense that for certain homogeneous risk groups the assumption that the $C_{i,j+1}/C_{i,j}$

[3] Asbestos-related diseases have particularly long latency periods, sometimes in excess of 40 years. See [21].

is log-normally distributed is empirically reasonable. However, in practice, each group of risks should be investigated critically. Actuaries prefer to investigate the ratio of subsequent cumulative or incremental claims, hence the convention of $C_{i,j+1}/C_{i,j}$ is followed in this section.

4.2.1 Log-normal model

The following assumptions are based on [111]. Triangular elements $C_{i,j}$ $(i, j \in \mathbb{Z}_+,\ i + j \leq I + 1)$ indicate cumulative claims. The so-called $F_{i,j} = \frac{C_{i,j}}{C_{i,j-1}}$ development factors are log-normally distributed random variables, with μ_j log-scale and σ_j^2 shape parameters, with the constraint $C_{i,0} = 1$. Hence, the logarithm of $F_{i,j}$ is governed by normal distribution with parameters μ_j and σ_j^2. Suppose that rows are independent. The distribution does not depend on i, which practically means that the reporting run-off of claims incurred in calendar year 2015 is independent of and identically distributed to the pattern of year 2014. In other words, F_{i,j_1} is independent of F_{k,j_2}, where $i \neq k$ and j_1, j_2 are arbitrary between 1 and I (can be equal). The parameter estimation is given by the solution of equations 4.1 and 4.2, except for $\hat{\sigma_I} := 0$. For further details see [111].

$$\hat{\mu_j} = \frac{1}{I - j + 1} \sum_{i=1}^{I-j+1} \log\left(\frac{C_{i,j}}{C_{i,j-1}}\right) \qquad j \in \{1, \ldots, I\} \tag{4.1}$$

$$\hat{\sigma_j^2} = \frac{1}{I - j} \sum_{i=1}^{I-j+1} \left(\log\left(\frac{C_{i,j}}{C_{i,j-1}}\right) - \hat{\mu_j}\right)^2 \qquad j \in \{1, \ldots, I - 1\} \tag{4.2}$$

The conditional distribution of $C_{i,k}$ for $i + k > I + 1$ or equivalently, the non-observed value is the last observed value in the row multiplied by the $F_{i,j}$ variables: for $k > I + 1 - i$, let $C_{i,k}|C_{i,I+1-i} \sim C_{i,I+1-i} \cdot F_{i,I+2-i} \cdot \ldots \cdot F_{i,k}$. This is how the forecast depends on the observed claims, and in order to determine the distribution of the product expression, parameters have to be estimated based on past observations. This principle is generally true for all models described in this section.

Observe that the distributions of $C_{i,I}$ ultimate payment values related to claims occurred in year i are also log-normal with log-scale parameter $\sum\limits_{k=1}^{I} \mu_k$, and shape parameter $\sum\limits_{k=1}^{I} \sigma_k^2$. However, an explicit expression with respect to the distribution of $UC = \sum\limits_{i=1}^{I} C_{i,I}$ does not exist, given that this is the sum of I log-normal variables.

4.2.2 Negative binomial model

The recursive construction of the negative binomial development model below is based on [111]. Rows are assumed to be independent, and in each row, the distribution of $X_{i,j}$ increment value is Poisson$\Big(\Theta_{i,j-1} \cdot \big(\frac{\beta_j}{\beta_{j-1}} - 1\big)\Big)$. $\Theta_{i,j-1}$ stands for a $\Gamma(c_{i,j-1}, 1)$ random variable given $\{C_{i,j} = c_{i,j}\}$. In other words, the conditional distribution of $\Theta_{i,j-1}$ given $\{C_{i,j} = c_{i,j}\}$ is Gamma with shape parameter $c_{i,j}$ and rate parameter 1. Intuitively, every subsequent increment $X_{i,j}$ in one row is a Poisson random variable with frequency parameter $\Big(\Theta_{i,j-1} \cdot \big(\frac{\beta_j}{\beta_{j-1}} - 1\big)\Big)$, also referred to as mixing distribution. The larger the previous $C_{i,j}$ cumulative claim amount, the larger the Poisson expectation. The other component of the Poisson frequency is $\beta_j = \sum\limits_{k=1}^{j} \gamma_k$ for $j \in \{1, \ldots, I\}$, the partial sums of the $\underline{\gamma} = (\gamma_1, \ldots, \gamma_I)$ payout pattern. These γ_i values are the unknown parameters in the model, and the purpose of parameter estimation is to approximate them. The intuition behind the model is that γ_j determines the proportions of the reported claims to the total claims occurred in year i (rows), whilst the frequency of increment is also influenced by a random gamma mixing distribution, depending on the value of cumulative reported claims until year $i + j$. Let $f_j := \frac{\beta_j}{\beta_{j-1}}$ $(j = 1, \ldots, I-1)$, and suppose that $f_j > 1$, i.e. $\gamma_k > 0$ for all k, which means an increasing payout pattern. It can be shown that the conditional distribution of increment $X_{i,j}$ given $C_{i,j-1}$ is NegBinom$\Big(C_{i,j-1}, \frac{1}{f_{j-1}}\Big)$ $(j = 2, \ldots, I)$.

$X_{i,j}$ increments are conditionally independent. The payout pattern can be expressed as $\gamma_1 = p_1 p_2 \ldots p_{I-1}$, $\gamma_i = (1-p_{i-1})p_i \ldots p_{I-1}$ $(i = 2, \ldots, I-1)$, $\gamma_I = 1-p_{I-1}$, where p_is stand for the second parameters of the negative binomial distributions. The other way around, $p_i = \frac{\sum_{k=1}^{i} \gamma_k}{\sum_{k=1}^{i+1} \gamma_k}$, $i = 1, \ldots, I-1$. Let the parameter estimation be based on the maximum likelihood estimator with respect to parameters $p_1, \ldots, p_{I-1}$. Let X denote the upper triangle and suppose the first column to be fixed. The likelihood function is as follows.

$$L(\underline{p}, X) = \prod_{i=1}^{I-1} P_{\underline{p}}(row_i) = \prod_{i=1}^{I-1} P_{\underline{p}}(X_{i,2} = k_{i,2}, \ldots, X_{i,I-i+1} = k_{i,I-i+1}) =$$

$$\prod_{i=1}^{I-1} P_{\underline{p}}(X_{i,I-i+1} = k_{i,I-i+1} | X_{i,I-i} = k_{i,I-i}, \ldots, X_{i,2} = k_{i,2}, X_{i,1} = k_{i,1}) \cdot \ldots$$

$$\ldots \cdot P_{\underline{p}}(X_{i,2} = k_{i,2} | X_{i,1} = k_{i,1}) =$$

$$\prod_{i=1}^{I-1} \binom{C_{i,I-i} + k_{i,I-i+1} - 1}{k_{i,I-i+1}} p_{I-i}^{C_{i,I-i}} (1-p_{I-i})^{k_{i,I-i+1}} \cdot \ldots \cdot \binom{C_{i,1} + k_{i,2} - 1}{k_{i,2}} p_1^{C_{i,1}} (1-p_1)^{k_{i,2}} =$$

$$\prod_{j=1}^{I-1}\prod_{i=1}^{I-j}\binom{C_{i,j}+k_{i,j+1}-1}{k_{i,j+1}}p_j^{C_{i,j}}(1-p_j)^{k_{i,j+1}} =$$

$$const\cdot\prod_{j=1}^{I-1} p_j^{\sum\limits_{i=1}^{I-j} C_{i,j}}(1-p_j)^{\sum\limits_{i=1}^{I-j} k_{i,j+1}} \longrightarrow \max_{\underline{p}}.$$

Let $l(\underline{p}, X)$ be $\log L(\underline{p}, X)$, i.e. the log-likelihood is

$$l(\underline{p}, X) = const + \sum_{j=1}^{I-1}\sum_{i=1}^{I-j} C_{i,j}\log p_j + \sum_{j=1}^{I-1}\sum_{i=1}^{I-j} k_{i,j+1}\log(1-p_j), \tag{4.3}$$

that infers the following system of equations:

$$\frac{\partial}{\partial p_j} l = \frac{\sum\limits_{i=1}^{I-j} C_{i,j}}{p_j} - \frac{\sum\limits_{i=1}^{I-j} k_{i,j+1}}{1-p_j} = 0, \tag{4.4}$$

$j \in \{1, \ldots, I-1\}$. Suppose that $\forall j \sum\limits_{i=1}^{I-j} k_{i,j+1} > 0$. This gives the estimators

$$\hat{p}_j = \frac{\sum\limits_{i=1}^{I-j} C_{i,j}}{\sum\limits_{i=1}^{I-j} C_{i,j+1}}, \tag{4.5}$$

which relate to the chain ladder development factors.

In [35] another approach of the Negative Binomial Model can be found. In this paper a dispersion parameter is included, and increments are so-called overdispersed negative binomially distributed random variables with mean $(f_j - 1)\cdot C_{i,j-1}$ and variance $\phi f_j(f_j-1)\cdot C_{i,j-1}$.

4.2.3 Poisson model

The Poisson model is described extensively in [111]. Suppose that the increments are independent $X_{ij} \sim Poisson(\mu_i\gamma_j)$ variables, $i, j = 1, \ldots, I$. In contrast to the previous two models, the Poisson frequency here depends on the year of original claim occurrence through μ_i. The estimation of $\gamma_1, \ldots, \gamma_I$ payout pattern values works exactly the same way as in the Negative Binomial case. Suppose that the upper triangle is given, and used in probabilities as a condition, denoted by $\mathcal{D}$. The estimation of the μ_i parameters is given by row-wise approximation, assuming the ratio of cumulative claims and payout pattern ratio until $I-i+1$:

$$\hat{\mu}_i = \frac{\sum\limits_{k=1}^{I-i+1} X_{ik}}{\sum\limits_{k=1}^{I-i+1} \hat{\gamma}_k}. \tag{4.6}$$

Since one of our aims is to generate ultimate claim values, assuming that the real parameters are $\hat{\gamma}_i$ and $\hat{\mu}_i$, one random ultimate claim variable is the sum of upper triangle values and Y lower triangle: $UC = \sum\limits_{i=1}^{I} C_{i,I-i+1} + Y$, where $Y \sim Poisson\left(\sum\limits_{i=2}^{I} \hat{\mu}_i(\hat{\gamma}_{I-i+2} + \ldots + \hat{\gamma}_I)\right)$.

4.2.4 Overdispersed Poisson model

A more general model frequently applied in practice is the Poisson model with a dispersion parameter.

Definition 4.4 (overdispersed Poisson distribution) Let $Y_{ij} \sim Poisson\left(\frac{\mu_i\gamma_j}{\phi}\right)$ and $\phi \in \mathbb{R}_+$. $X_{ij} := \phi Y_{ij}$ is called overdispersed Poisson random variable with dispersion parameter ϕ.

The motivation behind the definition is that the variance of some variables is directly proportional to the expectation, but with a proportionality factor other than 1. Observe that $E(X_{ij}) = \mu_i\gamma_j$ and $Var(X_{ij}) = \phi\mu_i\gamma_j$. Similarly to the Poisson model, $X_{i,j}$s are independent, although a new unknown parameter ϕ is included in addition.

In order to estimate $\underline{\mu}$ and $\underline{\gamma}$ parameters, the regular chain ladder method can be applied, resulting in unbiased estimation. Chain ladder is the most frequently used *deterministic* reserve calculation method in the actuarial practice, see [35]. In addition, evaluation of ϕ requires the determination of Pearson residuals, defined as

$$\hat{r}_{ij}^P = \frac{x_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}} = \frac{x_{ij} - \hat{\mu}_i\hat{\gamma}_j}{\sqrt{\hat{\mu}_i\hat{\gamma}_j}} \tag{4.7}$$

in the Poisson case. Generally, the Pearson residual stands for the standardised

$$\frac{\text{observation - expected value}}{\text{standard deviation}}.$$

Thus

$$\hat{\phi}_P = \frac{\sum\limits_{i+j \le I+1} \left(\hat{r}_{ij}^P\right)^2}{N-p}, \tag{4.8}$$

where $N = \frac{I(I+1)}{2}$ denotes the number of observations and $p = 2I - 1$ is the number of predicted unknown parameters. This method provides a biased estimator for the ϕ and also for the $\underline{\mu}$ parameters, see [111, 34].

The distribution of UC, similarly to the Poisson model is $UC = \sum_{i=1}^{I} C_{i,I-i+1} + Y$, where $Y \sim \phi \cdot Poisson\left(\frac{1}{\phi}\sum_{i=2}^{I}\hat{\mu}_i(\hat{\gamma}_{I-i+2} + \ldots + \hat{\gamma}_I)\right)$

Another possible parameterisation of the model is with the real parameters $\alpha_1, \ldots, \alpha_I; \beta_1, \ldots, \beta_I; c; \phi$. Incremental claims in turn are defined as $X_{ij} \sim \phi \cdot Poisson\left(\frac{e^{\alpha_i+\beta_j+c}}{\phi}\right)$, therefore $E(X_{ij}) = e^{\alpha_i+\beta_j+c}$ and $Var(X_{ij}) = \phi e^{\alpha_i+\beta_j+c}$. Parameters α and β can be estimated using the maximum likelihood method under the usual constraint that $\alpha_1 = \beta_1 = 0$.

4.2.5 Gamma model

On the one hand, the model described in [35] is the following: increments are assumed to be Gamma distributed random variables, i.e. $X_{ij} \sim \Gamma(\alpha, \beta)$, with expected value $E(X_{ij}) = m_{ij}$ and variance $Var(X_{ij}) = \phi m_{ij}^2$, given that parameters are $\alpha = \frac{1}{\phi}$ and $\beta = \frac{1}{\phi m_{ij}}$.

On the other hand, in [111] X_{ij} increments are deterministic sums of (r_{ij}) independent Gamma random variables, i.e. $X_{ij} = \sum_{k=1}^{r_{ij}} X_{ij}^{(k)}$, where $X_{ij}^{(k)} \sim \Gamma(\nu, \frac{\nu}{m_{ij}})$. Assuming that the rate parameters are equal, the distribution of X_{ij} is $\Gamma(\nu r_{ij}, \frac{\nu}{m_{ij}})$, and $E(X_{ij}) = r_{ij}m_{ij}$, $Var(X_{ij}) = \frac{r_{ij}}{\nu}m_{ij}^2$.

If $r_{ij} = 1 \ \ \forall i, j$, the above mentioned two models are identical, and $\nu = \frac{1}{\phi}$. The only difference is that r_{ij}s can be arbitrary integers in the latter model, allowing for higher degree of flexibility. In order to be able to perform estimations, knowing the upper triangle figures is not sufficient as the triangle of r_{ij} numbers is needed, as well. There exist $\mu_1, \ldots, \mu_I$ and $\gamma_1, \ldots, \gamma_I$ parameters under the constraint that $\sum_{i=1}^{I} \gamma_i = 1$, such that $E(X_{ij}) = r_{ij}m_{ij} = \mu_i\gamma_j$. The unknown $\mu_1, \ldots, \mu_I$, $\gamma_i, \ldots, \gamma_I$ and ν parameter values have to be approximated (or the inverse of the latter $\nu = \frac{1}{\phi}$). The estimates $\hat{\mu}_i$ and $\hat{\gamma}_j$ are averages of the observations weighted by numbers r_{ij}. For more details see Model Assumptions 5.19. in [111].

Parameters $\hat{\mu}$ and $\hat{\gamma}$ are the solution of the following system of equations:

$$\hat{\mu}_i = \frac{\sum_{j=1}^{I+1-i} \frac{X_{ij}}{\hat{\gamma}_j}}{I + 1 - i} \quad \text{and} \quad \hat{\gamma}_j = \frac{\sum_{i=1}^{I+1-j} \frac{X_{ij}}{\hat{\mu}_i}}{I + 1 - j}. \tag{4.9}$$

For technical reasons, the simple chain ladder method has been implemented instead of solving the above mentioned system of equations. The estimation of parameter ν using Pearson residuals is the following: for $i+j \leq I+1$ let

$$\hat{r}_{ij}^P = \frac{x_{ij} - \hat{\mu}_i \hat{\gamma}_j}{\hat{\mu}_i \hat{\gamma}_j}, \tag{4.10}$$

thus

$$\hat{\phi}_P = \frac{\sum\limits_{i+j \leq I+1} \left(\hat{r}_{ij}^P\right)^2}{\frac{I(I+1)}{2} - (2I-1)}, \tag{4.11}$$

resulting in $\hat{\nu}_P = \frac{1}{\hat{\phi}_P}$ as the estimator for ν.

4.3 Data and simulation of scenarios

Simulations have been performed on the basis of parameters derived from 5 run-off triangles, implemented in `R`. All sample triangles stem from disclosed industrial data:

1. RAA is an accumulated claims triangle from the Automatic Facultative business in General Liability, originally published in 1991 in Historical Loss Development, Reinsurance Association of America (RAA), also see in [35] and [69].
2. ABC is the run-off triangle of a workers' compensation portfolio of a large company, see [9], including an in-depth analysis of the data array.
3. GenIns is a general claims data triangle from [103].
4. M3IR5 is a simulated triangle from [113].
5. The fifth dataset consists of claims payment data of a Hungarian insurance company presented in a record by record format [4]. It contains the 11-year data of 6 accident years, embracing $43,081$ claims and $44,735$ payments.

Triangles RAA, ABC, GenIns and the one derived from the detailed dataset can be found in Appendix C. On the basis of the run-off triangles (1-4) and the detailed dataset (5), scenarios have been generated, as described later in section 4.6. (Especially in the last case, random simulation cannot be omitted when approximating the ultimate claim distribution.)

Table 4–1 and Figure 4–1 show the estimated parameter values for 4 different run-off triangles, under the assumption that the underlying model distribution is Gamma (subsection 4.2.5). To explore classical chain ladder parameters see Table 4–2 and Figure 4–2.

We emphasise that the statistical analysis of these triangles is not in scope of the dissertation. This has already been done in the referred papers. However, in order to *set the simulation parameters* appropriately in the sense that they reflect reasonable values, it is necessary to approximate them from real data. The dissimilarity of triangles ensures different parameter sets.

	RAA	GenIns	M3IR5	ABC
μ_1	21,048	4,037,004	937,099	741,980
μ_2	17,507	5,491,430	1,078,656	884,863
μ_3	23,723	5,325,069	1,308,673	1,042,456
μ_4	29,562	4,792,225	1,466,318	1,076,123
μ_5	25,751	5,149,424	1,779,449	1,105,213
μ_6	18,680	5,323,949	2,101,738	1,132,621
μ_7	15,676	5,764,152	2,595,485	1,134,290
μ_8	22,141	6,448,198	2,910,728	1,428,005
μ_9	19,019	5,582,284	3,227,487	1,938,165
μ_{10}	18,402	4,969,824	3,986,625	2,362,602
μ_{11}			4,286,695	2,688,976
μ_{12}			5,351,440	
μ_{13}			5,372,112	
μ_{14}			5,781,157	

	RAA	GenIns	M3IR5	ABC
γ_1	0.112	0.069	0.100	0.185
γ_2	0.224	0.172	0.094	0.241
γ_3	0.210	0.181	0.089	0.179
γ_4	0.148	0.193	0.083	0.121
γ_5	0.119	0.107	0.075	0.082
γ_6	0.092	0.075	0.078	0.059
γ_7	0.038	0.069	0.070	0.041
γ_8	0.031	0.047	0.068	0.031
γ_9	0.017	0.070	0.066	0.024
γ_{10}	0.009	0.017	0.067	0.019
γ_{11}			0.060	0.016
γ_{12}			0.062	
γ_{13}			0.046	
γ_{14}			0.042	

	RAA	GenIns	M3IR5	ABC
ν	2.22	9.25	80.78	137.79

Table 4–1: Parameters – Gamma Model.

	RAA	GenIns	M3IR5	ABC
λ_2	2.999	3.491	1.936	2.309
λ_3	1.624	1.747	1.458	1.421
λ_4	1.271	1.457	1.292	1.200
λ_5	1.172	1.174	1.205	1.113
λ_6	1.113	1.104	1.177	1.073
λ_7	1.042	1.086	1.136	1.048
λ_8	1.033	1.054	1.115	1.034
λ_9	1.017	1.077	1.101	1.026
λ_{10}	1.009	1.018	1.092	1.020
λ_{11}			1.076	1.016
λ_{12}			1.073	
λ_{13}			1.050	
λ_{14}			1.044	

Table 4–2: Chain ladder development factors – $\lambda_2, \ldots, \lambda_n$.

4.4 Stochastic claims reserving

4.4.1 Parametric models

Recall the parametric models in section 4.2. Firstly, parameters have been estimated in line with the upper triangles. Secondly, the conditional distributions of

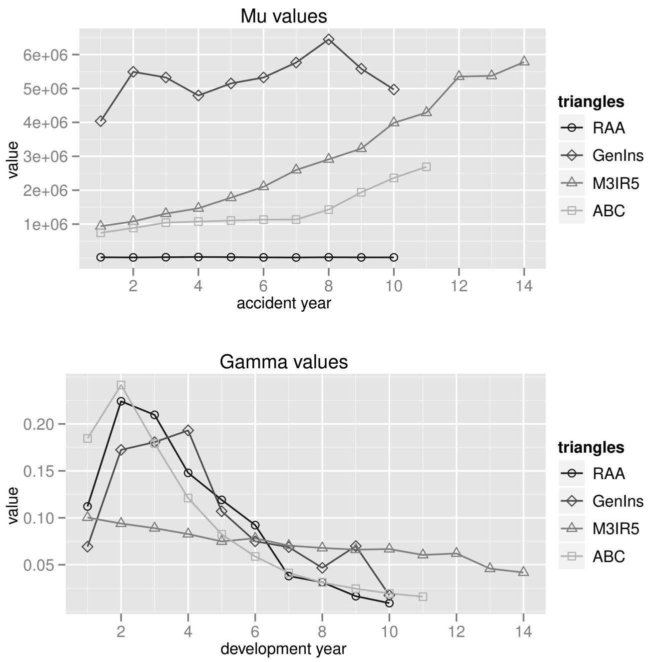


Figure 4–1: Parameters – Gamma Model.

the lower triangles have been approximated by determining the conditional distribution using the estimated parameters. Certainly, the approach has its drawbacks, producing inaccurate prediction intervals.

A *prediction interval* of an estimation is an interval where the predicted random value falls with a probability of $1-\alpha$. We choose the $1-\alpha$ parameter to be either 66% or 90%. Observe that this concept is different than the *confidence interval*, which provides an interval with respect to an unknown parameter. One may apply the calculation of the prediction interval for the $(n+1)$th member of an i.i.d. (independent and identically distributed) sample of normal distribution from the first n members. As the conditional distribution of the lower triangle only exceptionally has an analytical form, Monte Carlo type methods have been applied by generating 5000 lower triangles.

4.4.2 Bootstrap methods with overdispersed Poisson and gamma distributions

Bootstrapping in the mathematical sense has a dedicated literature and has been studied for almost four decades, well before applications in insurance emerged. The original introduction dates back to [32, 31] as a generalisation of jackknife, enhancing the power of available sample by resampling. Introducing an application of

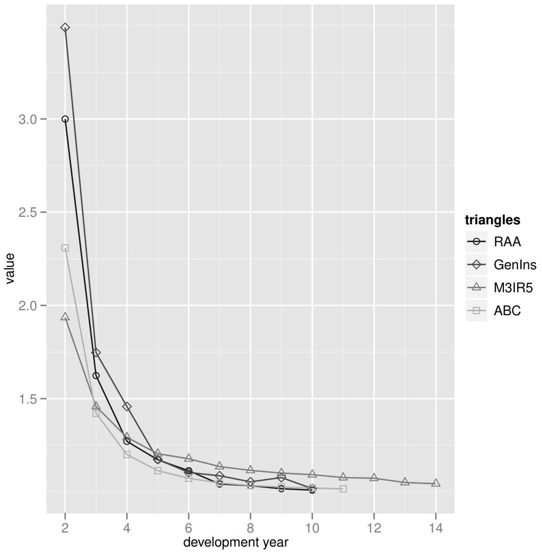


Figure 4–2: Chain ladder development factors.

bootstrapping in insurance, [5] was among the first papers, estimating distribution error. Later, [33] analysed the prediction error in conjunction with generalised linear models (GLMs) with bootstrapping, whilst [85] proposed an alternative bootstrap procedure to the previous one, using corrected residuals. The capability of error prediction was the basis of the concept which had driven the development of such models in the actuarial field. Contrary to the simple chain ladder model, it allows for capturing the variability of the outcome. More recent achievements are [11, 64] and a more practical guide is [94]. Thus, in recent years models using bootstrapping have become widely applied in actuarial practice, and have been studied in numerous works. For a more comprehensive overview the reader is advised to see [35, 111].

Let $X_{i,j}$ denote the incremental claim pertaining to accident year i and development year j. The core of the bootstrap models is the underlying GLM defined by

$$E(X_{i,j}) = \hat{X}_{i,j} \text{ and } Var(X_{i,j}) = \phi \hat{X}^{\rho}_{i,j}, \quad i = 1, \dots, I, \; j = 1, \dots, J, \tag{4.12}$$

where

$$\hat{X}_{i,j} = \exp(const + \alpha_i + \beta_j) \text{ and } \alpha_1 = \beta_1 = 0, \tag{4.13}$$

see [33]. Value ρ is the power of the variance function, which is $\rho = 1$ in the overdispersed Poisson and $\rho = 2$ in the Gamma case. In practice, instead of bootstrapping the original data, we bootstrap the residuals. From the various residual definitions such as the Pearson, the Deviance or the Anscombe, here we apply the first one, defined as the normalised sample $r_{i,j}^{(P)} = \frac{X_{i,j}-\hat{X}_{i,j}}{\sqrt{\hat{X}_{i,j}}}$. Having calculated the residuals, a random sample is drawn with replacement. The bootstrapped data sample is calculated by solving the previous equation backwards, i.e. $X_{i,j}^{*} = r_{i,j}^{(P)^{*}}\sqrt{\hat{X}_{i,j}} + \hat{X}_{i,j}$. In addition, the scale parameter ϕ is estimated from the residuals and the degrees of freedom. Repeating the process several times results in multiple sample triangles and hence, a distribution of ultimate claims. Algorithmically, the bootstrap models for claims reserving consist of two main general steps: parameter estimation to obtain the adjusted Pearson residuals and scale parameter, and an iterative process simulating future claims, see also [4].

(Step A) First step with parameter estimation, obtaining adjusted Pearson residuals:

1. Use the cumulative upper triangle to get development factors $\lambda_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}$ $(j = 1, \ldots, J-1)$.

2. Starting from the (anti-)diagonal element $C_{i,I-i+1}$, calculate backwards the cumulative elements $\hat{C}_{i,k} = \frac{1}{\lambda_{k+1}\ldots\lambda_{I-i+1}} C_{i,I-i+1}$. Perform this sub-step for each row.

3. As $\hat{C}_{i,j}$s determine the fitted $\hat{X}_{i,j}$ increments in the upper triangle $(i+j \leq I+1)$ by taking subsequent differences. Let $r_{i,j}^{(P)} = \frac{X_{i,j}-\hat{X}_{i,j}}{\sqrt{\hat{X}_{i,j}}}$ denote the unscaled Pearson residuals, with $X_{i,j}$ original increments in the numerator.

4. Let $\hat{\phi}_P = \frac{\sum_{i+j \leq I+1} \left(\hat{r}_{ij}^{P}\right)^2}{\binom{I+1}{2}-(2I-1)}$ stand for the Pearson scale parameter, where the denominator contains the degrees of freedom as the difference between the number of observations and estimated parameters.

5. Let $r_{i,j}^{adj} = \sqrt{\frac{\binom{I+1}{2}}{\binom{I+1}{2}-(2I-1)}}\hat{r}_{ij}^{P}$ be the adjusted Pearson residuals for bias correction.

(Step B) Second, iterative step, the following is replicated for a sufficient amount of times (In the simulations presented in section 4.6, Step B is repeated 5000 times.):

1. From the $\binom{I+1}{2}$-element set of adjusted residuals get an equal sized sample with replacement. Let them be $r_{i,j}^{*},\ i+j \leq I+1$.

2. Solve each equation $r_{i,j}^{*} = \frac{X_{i,j}-\hat{X}_{i,j}}{\sqrt{\hat{X}_{i,j}}}$ for $X_{i,j}$, and interpret this as an alternative past. This implies a new cumulative upper triangle, where the classical chain ladder method is applied again in order to complete the triangle with the lower part $\hat{C}_{i,j}$, where $i+j > I+1$.

3. This results in $\hat{X}_{i,j}$ $i+j>I+1$ increments, which will be used as the mean of the process distribution (e.g. overdispersed Poisson, Gamma, etc.), and $\phi\hat{X}_{i,j}$ as the variance. Simulate payment increments to obtain an ultimate claim.

4.4.3 Semi-stochastic methods

A family of models with the idea that the chain ladder factors are bootstrapped directly is presented in [36]. Suppose that each subsequent cumulative claim has a multiplicative link to the previous one in accident year j through a random variable α_j. Furthermore, let α_js be mutually independent and governed by the discrete uniform distribution on the set

$$\left\{\alpha_j(i) = \frac{C_{i,j+1}}{C_{i,j}} : i = 1, \ldots, I-j\right\}.$$

The expectation of each α_j is $E(\alpha_j) = \frac{1}{I-j}\sum\limits_{i=1}^{I-j}\frac{C_{i,j+1}}{C_{i,j}}$, equal to the average of the set of $\alpha_j(i)$ values, making the model of the type 'link ratios with simple average' method. In other words, this base model creates alternative lower triangles by completing each row recursively, choosing from $\alpha_j(i) = \frac{C_{i,j+1}}{C_{i,j}}$ factors randomly and processing $C_{i,j+1} = C_{i,j}\cdot\alpha_j$ for $i+j \geq I+1$. The ultimate claims are also random variables with parameters implied by the previous random recursion.

The first method, labeled as 'Uniform', is constructed to simulate a sufficiently large amount of lower triangles. Having simulated 5000 lower triangles, and calculated for each one the ultimate claim value, these total claim values determine the empirical distribution as a predictive distribution. The second method denoted as 'Unifnorm' is based on the assumption that the ultimate claim amount is a nearly normal random variable. The expectation of the ultimate claim is

$$E(UC) = \sum_{j=1}^{I} C_{I-j+1,j}\prod_{k=j}^{I-1} E(\alpha_k), \tag{4.14}$$

whilst the variance of the ultimate claim is

$$Var(UC) = \sum_{j=1}^{I}\left(C_{I-j+1,j}\left(\prod_{k=j}^{I-1} E\left(\alpha_k^2\right) - \prod_{k=j}^{I-1} E^2\alpha_k\right)\right). \tag{4.15}$$

4.5 Scores, errors, rankings

As noted in the previous section, stochastic reserving essentially predicts the distribution of the future payments, in contrast to traditional reserving, which purely results in a point estimation, see the chain ladder method. Hence, the result allows actuaries to analyse the volume of possible extreme outcomes and fit prediction intervals.

The probabilistic forecast as distribution dates back at least to [23], introducing the *prequential* principle. The term stems from the words *probabilistic forecasting with sequential prediction*, which refers to accumulating new observations from time to time, and implementing them into the subsequent days' estimations. A game-theoretic interpretation of probabilistic forecasts in the context of meteorological applications (similarly to the previous one) is analysed in [47], guiding readers through the predicting performance of a set of climatological experts. Observe the analogy between climate forecast experts and competing reserving methods. Both of these articles have a wide range of applicability going beyond meteorology, selecting the better performers from several rival models. [29] describes density forecast evaluation in a financial framework, including an example on probability integral transform on real S&P500 return data. Purely from a conceptual perspective, market data between '62 and '78 are in-sample, whilst the ones between '78 and '95 are out-of-sample observations. The set is split into these two parts in order to perform both a model estimation and an evaluation of the forecast. Drawing parallels between this financial example and our claims reserving task, the in-sample can be considered as the upper and the out-of-sample as the lower run-off triangle.

In other words, when the insurer decides to involve all past claim observations for the purpose of IBNR reserving, the figures by definition build up an upper triangle. For this reason, the usage of total quadrangles may seem to be counter-intuitive. However, in a longer run, the missing entries are filled and can be used for backtesting. New rows are unavoidably born at the same time with deficient elements on the right hand side of the row, which does not alter the fact that the older upper triangle is completed with a lower one. Depending on the total run-off period of the claims of a product, definite values become visible after 5 to 40 years, with the important discrepancy between the duration of fire (short) or liability (long) claims. Regulators of the insurance practice tend to use complete claim data sets available to them, which can also result in the truncation of a large triangle on its south-west and north-east part, where the north-west part tends to 0 for the reason of run-off. In contrast to liability insurance with potentially long hiding payout periods, in property insurance such as motor vehicle or homeowners insurance the run-off is not more than 3-4 years, allowing for a full quadrangle within 7 years of experience. Insurance companies do not usually have more triangles, apart from arranging the observations according to homogeneous risk groups, whilst regulators or oversight organisations do, see the example of NAIC in chapter 5. In the latter case it is of collective interest to use the triangles for the benefit of the participating insurance institutions. In addition, [4] proposes a simulation-based technique to complete lower triangles, particularly for heavy-tailed risk groups.

There is hardly any manner of ranking two forecasts in a way that all actuaries would agree with. Certainly, in case the predicting distribution coincides with the real distribution governing the sample, that one is the preference above all. Provided that in real life modelling questions professionals lack this exact knowledge, it is justifiable to create a ranking framework, which does not only take into account the mean square error of the prediction, but also other features discussed in the coming subsections. It is essential to understand how to assess these measures on the basis of available data and how to build a decision making framework in an algorithmic manner. Section 4.6 will outline the algorithmic steps of the simulation-based comparison process.

4.5.1 Probability integral transform

Probability integral transform (PIT) can be traced back to the early papers [83, 82] published consecutively by father and son from the Pearson family, as well as to the short remarks of [90] on multidimensional transformation. Later, the concept emerges in [23, 29, 47]. Statistical tests such as Kolmogorov–Smirnov or Cramér–von Mises decide whether or not to reject a certain distribution, however, they are deficient in suggesting what goes wrong with the hypothesis.

Suppose that an observation x_i is governed by an absolutely continuous distribution F_i, or density function f_i. Placing the observation into the argument of its own distribution function converts to a standard uniform random variable $F_i(x_i) \sim U(0,1)$ or $\int_{-\infty}^{x_i} f_i(u)du \sim U(0,1)$. Either be it one-dimensional or in higher dimension, this property will always be valid, except that in the latter case transformation has to be carried out with conditional distributions on the previous coordinates, see [90]. Now let $\hat{F}_i$ be the prediction given for F_i. Coinciding with the real distribution, F_i has a *necessary* condition such that $\hat{F}_i(x_i) \sim U(0,1)$. In its analysis of ranking histograms [51] introduced an illustrative counterexample with biased prediction and uniform PIT at the same time, disproving the uniform property as a *satisfying* condition. In other words, the PIT values related to the estimation distribute uniformly on (0,1), although the estimation itself is biased. The paper highlights the possible fallacies and misinterpretations of qualities that the rank histogram ensembles may conceal.

Proceed to the implementation of the PIT concept into the claims reserving model framework. A certain set of companies related to one business line[4] has n claim history quadrangles. For instance, take the 132 institutions for workers' compensation in the real life calculations later in chapter 5. Fix a reserving model and perform the ultimate claim value estimation for each of the triangles, followed by the observation of actually occured claims from the lower triangles. The latter stand for the realisation from the real unknown distribution, where the value is practically unknown for future estimation, but known for past data enabling validation. The result is n pair of $\{\hat{F}_i, x_i\}$ values that determine an empirical density on $(0, 1)$, and hence, the histogram of the PIT values $\hat{F}_1(x_1), \hat{F}_2(x_2), \ldots, \hat{F}_n(x_n)$. The predicted distribution functions of ultimate claims have been calculated with upper triangles fixed and the *real* occurred ultimate claim values have been substituted into them. This method is applicable if the amount of data is fairly large. Should the set consist of an extremely low number of data points, then the application of a randomised PIT or a non-randomised uniform version of PIT is more proper, see [22].

Comparing the histogram of PIT values with respect to its shape to the uniform density function is a way of evaluating the biasedness of a prediction. Generally, the deviation of the PIT histogram from uniformity reflects the dispersion of the predictive model. A ∩-shaped histogram can be translated as an overdispersed prediction with excessively wide prediction interval, i.e. overly heavy tailed distribution. Skewed histograms occur when central tendencies are biased. The variability of the data exceeds the fitted statistical model's variability. By contrast, ∪-shaped PIT suggests that the prediction shall be underdispersed with a narrow prediction interval, i.e. lighter tail than the underlying distribution would imply. In the latter case, the variability of the real governing distribution exceeds the variability of the model, whilst it is the other way around in the former case. Going forward, real-life observations and models result in histograms of less exemplary shapes, which are combinations of the mentioned two instances: skewed ∩-shaped PIT or a shape entirely biased towards 0 (or 1), for instance.

An equivalent approach to assess the quality of prediction from the perspective of its bias is the comparison of the empirical distribution function (DF) of the

[4]It is often used as a synonym of homogeneous risk groups throughout the dissertation that reflect insurance events of similar nature, such as fire or motor third-party liability or workers' compensation.

corresponding PIT values to the identity function. These curves are referred to as P-P plots.

Definition 4.5 (P-P plot) Let ξ be a random variable, its predictive distribution F and let $s_p := \sup_z\{z : \hat{F}(z) \leq p\}$. The *P-P plot* function is $p \mapsto P(\xi < s_p)$ ($[0,1] \to [0,1]$).

There is a bijection between the two concepts: integrating the PIT histogram tends to the P-P plot under fairly general circumstances. It also follows that a slanted-S shaped P-P plot corresponds to a $\cap$-shaped PIT. If $Qu(\eta_{\cdot j}, p)$ stands for the p-quantile value of the empirical distribution function defined by $\eta_{1j}, \ldots, \eta_{Mj}$, the empirical P-P plot is $\frac{1}{N}\sum_{j=1}^{N} \chi_{\{\xi_j < Qu(\eta_{\cdot j}, p)\}}$ for $p \in (0,1)$.

4.5.2 Continuous ranked probability score and energy score

Scores support the verification of probabilistic forecasts based on distribution estimates and observed outcomes. Probability scores provide an applicable technique to measure the quality of predicted distributions as introduced in section 2.5. In spite of their application in other disciplines, to our knowledge, scores have been researched to a limited extent in peer-reviewed journals in the context of technical reserving in insurance. In [4], a simulation-based method is constructed for the selection from competing models. As an extension of regression models in non-life ratemaking to generalised additive models for location, scale, and shape (GAMLSS), [58] compares various models based on their score contributions. In this analysis, Brier score, logarithmic score, spherical score and deviance information criterion (DIC) is used for Poisson, zero-inflated Poisson and negative binomial assumptions, whilst CRPS is also calculated for three zero-adjusted models. Using a real-life dataset, [104] compares the overdispersed Poisson, gamma and log-normal models in the bootstrap framework and their residual adjustments using the Dawid-Sebastiani scoring rule (DSS). In modelling claim severities and frequencies in automobile insurance, [50] considers scores for model comparison, which either apply or exclude spatial and certain claim number components.

Different distributions or using insurance claims prediction terminology, the competing models of reserving are analogous to different forecasters. Given that these models may either result in discrete or in absolutely continuous predictive distributions, it is of high practical relevance to select an appropriate score functional which is flexible enough to work with both cases. The following scoring rule is more robust than the logarithmic or Brier scores, and requires practically no assumption with regards to the distribution observed, let it be either discrete or not.

Definition 4.6 (Continuous ranked probability score (CRPS))

$$CRPS(F, x) = -\int_{-\infty}^{\infty} \left(F(u) - \chi_{\{x \leq u\}}\right)^2 du,$$

where indicator function $\chi_{\{x \leq u\}}$ equals 1 if $x \leq u$ and 0 otherwise.

F is supposed to be the predictive distribution function and x to be the observation. Some of the articles define positive CRPS, however, here we will use its negative counterpart. CRPS is designed to handle the case of prediction of distribution functions, and has been chosen to measure goodness-of-fit due to its rather general applicability, when no particular underlying distributional or parametric features of the random sample is assumed. CRPS can be considered as a generalisation of the Brier score (BS); it is the integral of BS over the domain of all threshold values, see [52]. In other words, there is a direct connection between the CRPS and an event-no-event score. Vice versa, the concept of energy score (ES) can be thought of as the generalisation of CRPS.

Definition 4.7 (Energy score) $ES_\beta(F, x) = \frac{1}{2}E_F|X - X'|^\beta - E_F|X - x|^\beta$ with an arbitrary constant $\beta \in (0, 2)$. Let X and X' be independent copies from probability distribution F. For $\beta = 1$, $ES_\beta(F, x) = CRPS(F, x)$, see [100].

CRPS can be evaluated directly, or in case of $\beta \neq 1$, it is feasible to approximate the energy score by sampling from the empirical distribution. The latter is substantially more time-consuming because the necessary sample size has to be drawn in order to reach a satisfying accuracy. From a practical perspective, the magnitude of difference with regards to computational time elapsed is approximately 100.

4.5.3 Empirical coverage and average width

The intention of the following definition is to grasp the consistency between the probability of falling out of a given interval assuming a predictive distribution, and the real distribution. In other words, the aim is to find the likelihood that a random variable of measure Q coincides with a central prediction interval determined by F. Meteorology related discussion can be found in [8]. For an application from the financial sector see [19], addressing conditional interval forecasts and asymmetric intervals, whilst the closest one to stochastic claims reserving can be found in [2, 4]. Both on coverage and average width the most detailed study is believably provided by

[47].

Definition 4.8 (Coverage α) Let Q stand for the probability measure governing the real distribution of the ultimate claim, and F for the forecast distribution. $Q\left(F^{-1}\left(\frac{1-\alpha}{2}\right), F^{-1}\left(\frac{1+\alpha}{2}\right)\right)$ is the central α prediction interval of F given Q.

The definition above results in the proportion of observations coinciding with the central prediction interval bounded by the lower and upper quantiles of the predictive distribution. In order to give the concept meaning in the context of run-off triangles and ultimate claims, conditional distributions have to be defined, given the upper tringles. Suppose that $\mathcal{D}_j$ is an upper triangle associated with the jth company. Fix an arbitrary model discussed in section 4.2, to be applied on each triangle in order to forecast claims. Let $Q_{\eta_j|\mathcal{D}_j}$ stand for the ultimate claim distribution resulted by the chosen model given $\mathcal{D}_j$, whilst $Q_{\xi_j|\mathcal{D}_j}$ is the actual conditional distribution. With the previous notations, the definition of coverage converts into

$$P_{Q_{\xi_j|\mathcal{D}_j}}\left(Q^{-1}_{\eta_j|\mathcal{D}_j}\left(\frac{1-\alpha}{2}\right) < \xi_j < Q^{-1}_{\eta_j|\mathcal{D}_j}\left(\frac{1+\alpha}{2}\right)\right). \tag{4.16}$$

It is easy to see that if η_j has identical distribution to ξ_j, which means a perfect prediction, expression 4.16 equals to α for any $\alpha \in (0,1)$. Now assume that the model determines the predictive distribution given $\mathcal{D}_j$ in the form of a random sample $\eta_{1,j}, \ldots, \eta_{M,j}$ for $j \in \{1, \ldots, n\}$ and arbitrarily large positive integer M. Let $Qu(\eta_{\bullet j}, p)$ stand for the p-quantile of the empirical distribution determined by sample $\eta_{1,j}, \ldots, \eta_{M,j}$. For $\alpha \in (0,1)$ the central prediction interval's approximation is $\frac{1}{n}\sum_{j=1}^{n}\chi_{\{Qu(\eta_{\bullet j}, \frac{1-\alpha}{2}) < \xi_j < Qu(\eta_{\bullet j}, \frac{1+\alpha}{2})\}}$, using χ_A for the notation of the indicator function of event A. That is given by generating a random sample of ultimate claims on the basis of the fixed model, conditionally on $\mathcal{D}_j$ for each $j \in \{1, \ldots, n\}$. In order to achieve convergence, increase the sample size M.

As an ancillary measure besides coverage, average width of prediction covers the expected difference between the lower and upper p-quantiles, a value expressed in actual payment. Alternatively, it is called the sharpness of the predictive evaluation. The narrower the width, the better the prediction.

Definition 4.9 (Average width (sharpness)) Let $Q_{\xi_j|\mathcal{D}_j}$ be the conditional probability measure of the ultimate claim based on a fixed model, provided that the upper triangle is $\mathcal{D}_j$. Suppose that there is an underlying multivariate distribution $Q_{\mathcal{D}}$ governing upper triangle $\mathcal{D}$. The average width of the model is

$$E_{Q_{\mathcal{D}}}\left(Q^{-1}_{\xi_j|\mathcal{D}_j}\left(\frac{1+\alpha}{2}\right) - Q^{-1}_{\xi_j|\mathcal{D}_j}\left(\frac{1-\alpha}{2}\right)|\mathcal{D}_j\right). \tag{4.17}$$

It is the expected difference between the upper and lower quantiles expressed in payments. This width can be interpreted as the prediction's sharpness. Similarly to the practical evaluation of coverage, generate for each upper triangle $\mathcal{D}_j$ a sufficiently large M amount of random ultimate claim values. Hence, the sharpness of the model given the set of run-off triangle observations is $\frac{1}{n}\sum_{i=1}^{n}\left(Qu(\eta_{\bullet j}, \frac{1+\alpha}{2}) - Qu(\eta_{\bullet j}, \frac{1-\alpha}{2})\right)$. ($Qu(\eta_{\bullet j}, x) = \eta^*_{\lfloor x\cdot M\rfloor, j}$ in the ordered sample.)

4.5.4 Mean square error of prediction

Measuring the expected squared distance between the predictor and the actual outcome has been part of the conventional way of actuarial reserving. We shall distinguish the conditional error given the upper triangle $\mathcal{D}$ and the unconditional one. Eventually, in the judgment of the specific model, the unconditional version is assessed in order to measure the average performance of the model without constraining it on a fixed run-off triangle. Several articles break down the definition on occurrence years, i.e. inspecting $C_{i,J}$ real and $\hat{C}_{i,J}$ estimated ultimate claims for occurrence year i, or the future (reserve) part of the claims $C_{i,J} - C_{i,J-i+1}$ real and $\hat{C}_{i,J} - C_{i,J-i+1}$. Without loss of generality, the definition in the present work is formalised for total ultimate claims $UC = \sum_{i=1}^{I} C_{i,J}$. For the sake of traceability, the definition contains the notation of $\xi_i \sim Q$ ultimate claim for company i and $\eta_i \sim F_i$ ultimate claim prediction.

Definition 4.10 (Mean square error of prediction (MSEP)) The conditional mean square error of prediction of estimator η_i for ξ_i given $\mathcal{D}_i$ is

$$msep_{\xi_i|\mathcal{D}_i}(\eta_i) = E\left((\xi_i - \eta_i)^2 \,|\mathcal{D}_i\right) \quad (i = 1, \ldots, n \text{ different triangles}). \tag{4.18}$$

The unconditional MSEP is

$$msep_{\xi_i}(\eta_i) = E\left((\xi_i - \eta_i)^2\right) = E\left(E\left((\xi_i - \eta_i)^2 \,|\mathcal{D}_i\right)\right). \tag{4.19}$$

The MSEP can be split into

$$E\left((\xi_i - \eta_i)^2|\mathcal{D}_i\right) = Var(\xi_i|\mathcal{D}_i) + (\eta_i - E(\xi_i|\mathcal{D}_i))^2, \tag{4.20}$$

where the first term is the variance of the process, whilst the second term reflects the estimation error. Similarly to the conditional version,

$$E\left((\xi_i - \eta_i)^2\right) = E\left(Var(\xi_i|\mathcal{D}_i)\right) + E(\eta_i - E(\xi_i|\mathcal{D}_i))^2. \tag{4.21}$$

Reserving models can be improved by minimising the second term. In conjunction with some parameteric models, MSEP can be derived in an analytical form. See [68] for the original Mack model and [14] in a time series method revisiting the result of the previous article.

Let $x_i := E(\xi_i|\mathcal{D}_i)$ and y_{ji} $j = 1, \ldots, M$, $i = 1, \ldots, n$ be the jth randomly generated ultimate claim based on an arbitrarily fitted model, given that the upper triangle is $\mathcal{D}_i$. Similarly to that, let z_{ji} stand for the jth ultimate claim scenario generated with the real parameters and development distribution, conditionally on upper triangle $\mathcal{D}_i$. Hence,

$$msep_{\xi_i|\mathcal{D}_i}(\eta_i) \approx \frac{\sum_{j=1}^{M}(z_{ji} - x_i)^2}{M-1} + \left(\frac{\sum_{j=1}^{M} y_{ji}}{M} - \frac{\sum_{j=1}^{M} z_{ji}}{M}\right)^2. \tag{4.22}$$

Calculating the MSEP values results in the pure $Var(\xi_i|\mathcal{D}_i)$ variance supposing the knowledge of real parameters instead of estimation, due to the fact that y_{ji}, z_{ji} are identically distributed.

4.5.5 Introductory examples for illustrative purposes

In order to present these methods in a transparent way, two examples are demonstrated here. The examples have not been made based on run-off triangle observations, but in a simplified manner for illustrative purposes. This means that no sample or observation is applied, it is instead assumed that the actuaries below could somehow calculate prediction. The first scenario has similarity to the one in [47].

Example 4.1 (continuous claim amount – Example 1) Suppose that a non-life insurance firm has already paid an amount of 1 for the accidents occurred in year 2013 in its automobile liability business. Let ξ be the total amount paid till the end of 2014 for accidents that occurred in 2013. Assume that ξ is of log-normal distribution with log-scale and shape parameters μ and 1, where μ is supposed to be a standard normal random variable.

actuary	predictive distribution
ideal	$LN(\mu, 1)$
long-term	$LN(0, 2)$
ordinary	$\frac{1}{2}LN(\mu, 1) + \frac{1}{2}LN(\mu + \delta, 1)$, where $\delta = \pm 1$ with probability $\frac{1}{2}, \frac{1}{2}$
intern	$LN(-\lvert\mu\rvert, \sigma^2)$, where $-\lvert\mu\rvert + \frac{\sigma^2}{2} = \mu + \frac{1}{2}$, i.e. $\sigma^2 = \begin{cases} 4\mu + 1 & \mu \geq 0 \\ 1 & \mu < 0 \end{cases}$

Table 4–3: Predictive distributions of the 4 actuaries regarding Example 1 (Log-normal). $\xi \sim LN(\mu, 1)$, where $\mu \sim N(0, 1)$.

actuary	predictive distribution
ideal	Poisson$(x \cdot \lambda)$
long-term	NegBinom$(1.5, \frac{1}{1+x \cdot 0.5})$
ordinary	$\frac{1}{2}$Poisson$(x \cdot \lambda) + \frac{1}{2}$Poisson$(x \cdot \lambda \cdot \delta)$, where $\delta = 1 \pm \frac{1}{10}$ with probability $\frac{1}{2}, \frac{1}{2}$
intern	NegBinom$(2x \cdot \lambda, \frac{2}{3})$

Table 4–4: Predictive distributions of the 4 actuaries regarding Example 2 (Poisson). $\eta \sim$Poisson$(x \cdot \lambda)$, where $\lambda \sim \Gamma(1.5, 0.5)$ and $x = 1000$.

Example 4.2 (discrete claim count – Example 2) In the integer valued example, the number of liability insurance claims for damages incurred and reported in 2013 was 1000. Let η be the number of claims for damages of 2013 and reported in 2014. Assume that the distribution of η is Poisson$(1000 \cdot \lambda)$, where λ is a gamma distributed random variable with shape parameter 1.5 and rate parameter 0.5.

Four imaginary actuaries have been compared based on their performance in predicting the distributions in the two examples. The *ideal actuary* knows all the relevant circumstances and he or she is aware of the exact value of μ in *Example 1*, and λ in *Example 2*. The *long-term actuary* does not intend to familiarise himself with the actual information and simply assesses the unconditional distribution. The *ordinary actuary* attempts to estimate the parameters, inherently including some possible error into the estimation. At last, the *intern actuary* finds the expected value without caring about the distribution itself. Table 4–3 and Table 4–4 give an overview of the distributions and the predicted distributions. Year 2014 has been simulated 10,000 times and the different probabilistic forecasts have been compared based on these.

4.5.5.1 Probability integral transform

PIT histograms with respect to the two examples can be seen in Figure 4–3. Dashed lines represent uniform distributions. These figures provide examples for inaccurate probabilistic forecasts in spite of their appropriate PIT histograms,

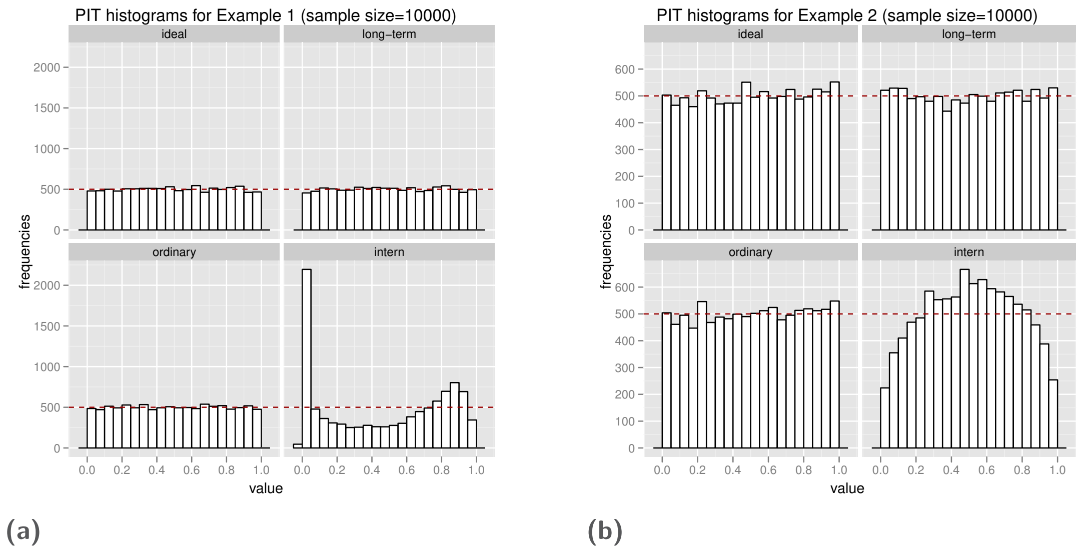


Figure 4–3: PIT histograms for (a) Example 1 (Log-normal case) and (b) Example 2 (Poisson case).

since the shapes of the long-term and of the ordinary predictions deviate not more considerably from uniform distribution than in the ideal case. Nevertheless, the intern case instantly demonstrates inappropriateness.

4.5.5.2 Continuous ranked probability score

Table 4–5 presents the score results in expectation. Consider the Log-normal example in case of the ideal actuary. Given $\{\mu = m\}$ and $\{\xi = x\}$, the score value is in the form $\int_0^\infty \left(\Phi(\log y - m) - 1_{\{y \geq x\}}\right)^2 dy$. Observe that the score measure provides proper ranking, however, the difference is not substantial between the ideal and the intern values with respect to the second example.

	ideal	long-term	ordinary	intern
Example 1	-0.48	-1.86	-0.63	-1.83
Example 2	-14.49	-327.62	-32.62	-14.64

Table 4–5: CRPS results in Example 1 and Example 2.

4.5.5.3 Coverage and average width

Having looked at Table 4–6 and Table 4–7, it can be observed that inappropriate distributions may provide good results. The coverage of the long-term and of the ordinary actuary, and the average width of the intern actuary is acceptable as well. However, in the two examples only the ideal actuary has presented suitable values for both measures.

Actuary	Coverage (%)		Average width	
	66%	90%	66%	90%
ideal	67.0	90.5	1.93	3.29
long-term	67.3	90.5	2.74	4.65
ordinary	67.2	90.4	5.01	11.64
intern	47.3	74.2	2.87	4.88

Table 4–6: Coverage and Average width in Example 1 (Log-normal).

Actuary	Coverage (%)		Average width	
	66%	90%	66%	90%
ideal	66.2	89.5	48.91	83.16
long-term	65.5	89.5	1052.00	1868.00
ordinary	66.4	89.5	98.24	140.62
intern	76.2	95.2	59.89	101.84

Table 4–7: Coverage and Average width in Example 2 (Poisson).

4.5.5.4 Mean square error of prediction

We elaborate on the msep of the four actuaries according to Example 2:

$$\begin{aligned} msep_\eta(\text{ideal}) = msep_\eta(\text{intern}) = E(msep_{\eta|\lambda}(x\lambda)) = \\ = E(Var(\eta|\lambda)) + E((x\lambda - E(\eta|\lambda))^2) = E(x\lambda) = x \cdot \alpha \cdot \beta \end{aligned} \tag{4.23}$$

$$\begin{aligned} msep_\eta(\text{long-term}) = msep_\eta(E(\eta)) = Var(\eta) = \\ = \frac{\alpha \left(1 - \frac{1}{1+x\cdot\beta}\right)}{\left(\frac{1}{1+x\cdot\beta}\right)^2} = x \cdot \alpha \cdot \beta \cdot (1 + x \cdot \beta) \end{aligned} \tag{4.24}$$

$$\begin{aligned} msep_\eta(\text{ordinary}) = E\left(msep_{\eta|\lambda}(\frac{1}{2} \cdot x\lambda \cdot (1+\delta))\right) = \\ = E(Var(\eta|\lambda)) + E((\frac{1}{2} \cdot x\lambda \cdot (1+\delta) - E(\eta|\lambda))^2) = \\ = x \cdot \alpha \cdot \beta + \frac{x^2}{400} \cdot E(\lambda^2) = x \cdot \alpha \cdot \beta \left(1 + \frac{x \cdot (1+\alpha) \cdot \beta}{400}\right) \end{aligned} \tag{4.25}$$

where $\alpha = 1.5$, $\beta = 0.5$ denote the shape and rate parameters of the gamma distributions. A weakness of the MSEP is that the intern is as good as the ideal one. Results in conjunction with the two examples are shown in Table 4–8. The tables show that in the long-term case, the prediction is inaccurate despite the PIT histogram and despite the fact that the prediction interval coverage reflected appropriate fit.

	ideal	long-term	ordinary	intern
Example 1	34.5	47.2	37.2	34.5
Example 2	750.0	375750.0	3093.8	750.0

Table 4–8: Mean Square Error of Prediction in Example 1 and Example 2.

4.6 Simulation and results

4.6.1 Preliminary remarks

Simulations have been implemented in R. Documentation and user manual regarding the self-developed codes, as well as the detailed simulation results for several parameter sets can be received on request, whilst the ChainLadder package's vignette can be found on the website of the r-project [43].

In this section, a Monte Carlo type method will be introduced, followed by simulations, consisting of corresponding goodness-of-fit values described in section 4.5. Parameters of the example come from the run-off triangle RAA, which is an accumulated claims triangle from the Automatic Facultative business in General Liability, originally published in Historical Loss Development, Reinsurance Association of America (RAA), 1991, and was also used as an example in [35] and [69]. Besides providing real life parameters via parameter estimation, RAA has no other impact on the calculations, i.e. several distributional models have been fitted to get the parameter values from real data. Similar calculations to the ones on RAA have been done for triangles ABC, GenIns and M3IR5, discussed in section 4.3. However, they have been excluded from the dissertation due to largely identical conclusions.

4.6.2 Monte Carlo type method

To perform the comparison of various claims reserving methods in case of different distributional backgrounds of run-off triangles, a Monte Carlo type method (MC method) has been constructed. The objective of this method is the establishment of a score-based ranking between the different stochastic claims reserving techniques for cases when the real development property of claims payments for accident years follows a specific model from section 4.2. Calculations have been implemented in line with the $\sum_{i=1}^{I} C_{i,I}$ ultimate claim values, i.e. the complete run-off of claim payments related to the set of accident years $1, \ldots, I$. However, this has been done without restriction of generality, and can be translated into $\sum_{j=2}^{I} X_{I-j+2,j}$ payments on a one-year time horizon, without any specific alteration. Furthermore, as emphasised in section 4.1, the actual payments deducted from the ultimate forecasted claims result in the insurance reserves. Note that insurance companies might be more interested in the one-year horizon perspective due to the stipulations of prospective Solvency II directives.

Algorithm 4.1 (Arató, Mályusz, Martinek) Reserving model evaluation (Figure 4–4).

step I | Given the development distribution model of run-off triangles and corresponding parameters, as a *first step*, N run-off triangles have to be generated independently. For each triangle the corresponding ultimate claim values are $UC_1, UC_2, \ldots, UC_N$. Let $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N$ denote the upper triangle conditions.

step II | In the *second step*, for each generated upper triangle, the predicted distribution of ultimate claims is evaluated taking into consideration the methods described in section 4.4. These predicted distributions are determined via the Monte Carlo method. In case of the parametric models, parameter values for each generated upper triangle have to be estimated, thus these will show discrepancy compared to the real parameters. Assuming these parameter values for each upper triangle condition ($\mathcal{D}_j$), M ultimate claim values have to be drawn, denoted by $\hat{UC}_{1,j}, \hat{UC}_{2,j}, \ldots, \hat{UC}_{M,j}$, as stochastic predictors. Collectively these result in the predictive distribution. Forecasts have to be made utilising the two bootstrap methods, the uniform method and the uniform normal method.

step III | At last, in the *third step*, for each pair UC_j and tuple $(\hat{UC}_{1,j}, \hat{UC}_{2,j}, \ldots, \hat{UC}_{M,j})$ scores, PITs, MSEP (and quantile values), coverage and average width have to be calculated according to section 4.5. The predictive distribution function F_j derives from values $\hat{UC}_{1,j}, \hat{UC}_{2,j}, \ldots, \hat{UC}_{M,j}$, and the corresponding c value is UC_j.

Results are shown on several figures, provided that the Gamma model (subsection 4.2.5) governs the claim run-off. Let $N := 2000$ and $M := 5000$. Figure 4–5 contains the histograms of 2000 PIT values. On the one hand, Figure 4–7 consists of boxplot

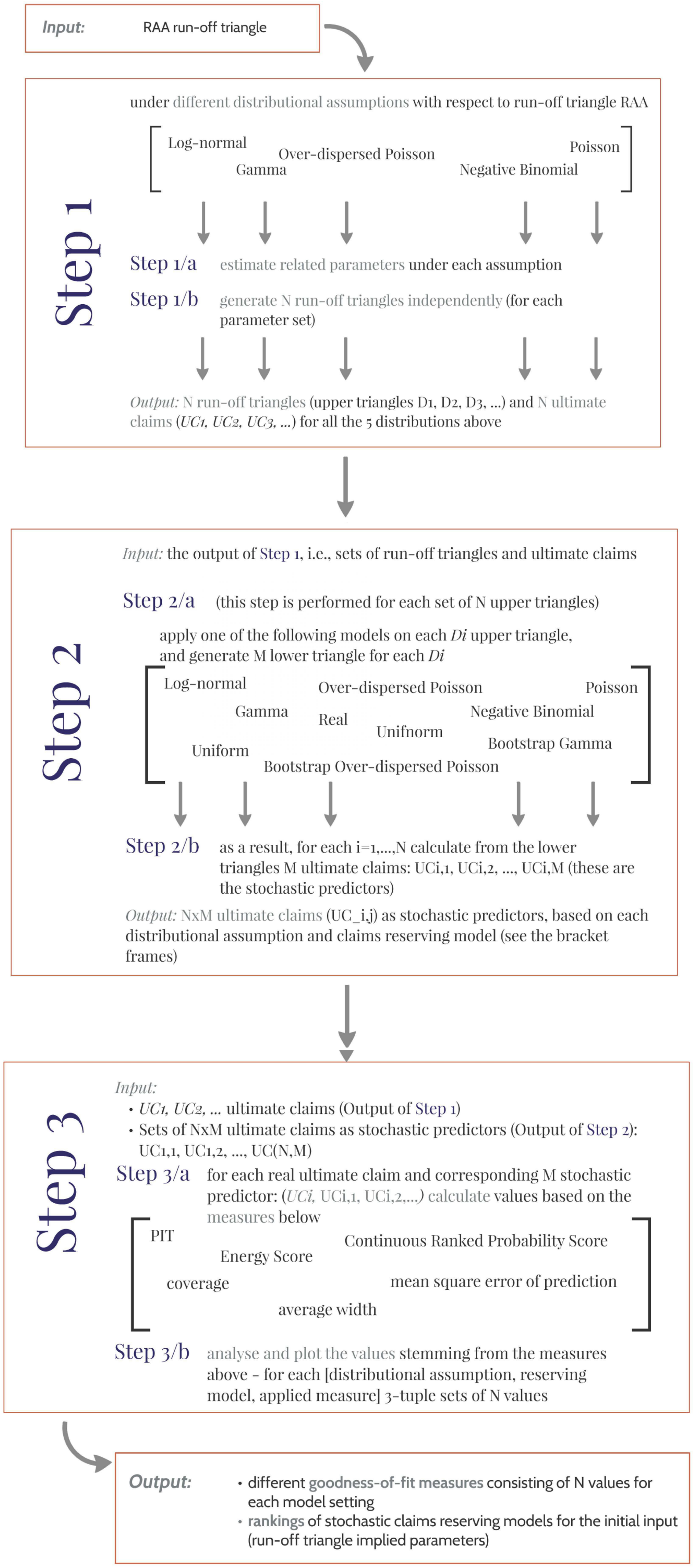


Figure 4–4: Scheme of the simulation-based algorithm (4.1).

representations of CRPS scores for each claims reserving technique. On the other hand, the mean CRPS values can be found in Table 4–9. Similarly, Figure 4–8 represents the energy scores for the $\beta = \frac{1}{2}$ setting. Recall that a higher score value means a better rank among reserving methods. The label 'Ideal' refers to an extreme forecast method exposed only to the stochastic volatility of the model, being aware of its underlying mathematical parameterisation. Hence, the 'Ideal' predictor knows the exact underlying model and its parameters, exploiting it in the second step, instead of estimation based on the upper triangle condition. Theoretically, the 'Ideal' predictor leads to the best possible manner to predict claims reserves in expectation, and is included for reference purposes, similarly to the ideal actuary in subsection 4.5.5.

4.6.3 Results on a gamma distributed world

Suppose that the claim run-off is in accordance with the gamma distribution model. Results can be interpreted in different ways, taking into consideration the following relevant questions.

1. Which score or error number is the most consistent, and how do they correlate with each other? Do the equipments applied in stochastic predictions choose the actual models better than regular measures, such as the mean square error of prediction?
2. Which non-parametric, distribution-free methods predict the distributions properly? Do they outperform prediction methods derived from parametric models?
3. How reliable and sharp are the prediction intervals?

Parameters stem from fitting the gamma model to the RAA data:

$$(\mu_1, \ldots, \mu_{10}) = (21048, 17507, 23723, 29562, 25751, 18680, 15676, 22141, 19019, 18402),$$

$$(\gamma_1, \ldots, \gamma_{10}) = (0.112, 0.224, 0.209, 0.147, 0.119, 0.092, 0.037, 0.031, 0.016, 0.009),$$

and $\nu = 2.22$. Based on the MSEP values in Table 4–9, the poor fit of the log-normal model is arguably conspicuous, whilst the other 4 parameter estimating models are relatively similar. The bootstrap methods slightly underperform, and the uniform and unifnorm methods provide rather poor results compared to the other ones.

As one might have expected PIT histograms in Figure 4–5 related to the Negative Binomial and to the Poisson distribution models demonstrate extremely poor fits. These provide shining examples for light-tailed predictions. The reason for this is that the model parameters imply for example a Poisson variable with a relatively high expected value, which makes the standard deviation very low compared to the

Res. Method	CRPS (mean)	En. Sc. (mean)	MSEP (mean)	MSEP (median)
Log-normal	-17840	-77.47	625.70	215.90
Negative Binomial	-9878	-81.03	17.99	9.86
Poisson	-10010	-84.23	17.99	9.86
Overdispersed P.	-7547	-57.10	18.00	9.91
Gamma	-6911	-54.28	15.39	9.11
Uniform	-20980	-80.25	614.40	471.60
Unif. Normal	-50040	-145.90	613.90	471.10
Bootstrap Gamma	-7856	-57.82	20.21	9.88
Bootstrap Od. Pois.	-7865	-57.84	20.15	10.05
Ideal	-4074	-41.53	5.30	5.30

Table 4–9: Scores and Mean Square Errors in the Gamma Model example (MSEP in units of 10e+07).

expectation. Hence, the ultimate claim values are in a very narrow range compared to the predicted distributions, leading to PIT histograms in which all occurrences are in a 10-20 percent wide probability range of the predicted distributions.

A further explanation of the behaviour with regards to the Poisson distribution is the following: Using a Poisson distribution as the predicted distribution, the difference between the smallest and the largest value of the empirical predicted distribution is very small, consequently, the real ultimate claim values will most of the time be either bigger or smaller than all the 5000 values of the predicted distribution. Therefore, the corresponding PIT graphs tend to contain occurrences almost exclusively in the 5 percent and 95 percent probability levels. In essence, the PIT analysis strongly suggests that the Poisson distribution should not be used as the incremental claims of the IBNR claims in the current example. The reason for this is that even though the MSEP values are acceptable, the low variance attribute leads to the Poisson distribution seemingly being of insignificant upgrade over a simple point estimation of the expected value. Nevertheless, the occurrence of nearly Poisson distributed triangles for payment amounts is not unprecedented on real data. Looking at the $\cap$-shaped histograms that derive from the bootstrap methods, we can see examples for heavy-tailed predictors.

During the examination of the P-P plots, the graph most similar to the identity function expected in the case when real distribution is of the same type as the one in the simulation method, means the best result. This has been validated by the calculations, see Figure 4–6. Despite the fact that the Gamma model fits correctly (to a triangle following the Gamma model), without being aware of the exact values of parameters, the prediction tends to be slightly overdispersed. The conclusion is that in terms of the consistency of P-P plot quality, the bootstrap methods provide acceptable

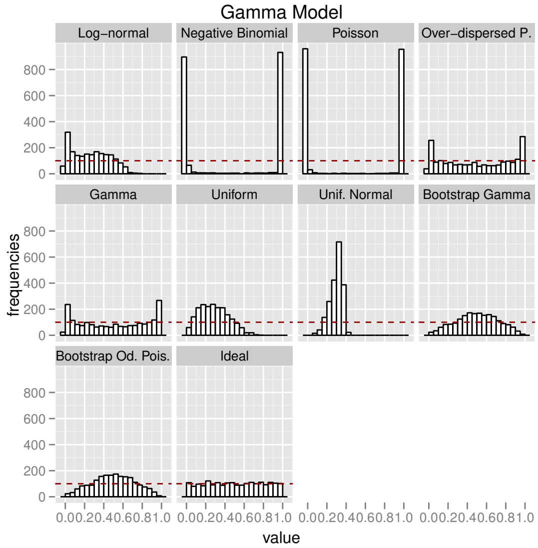


Figure 4–5: Histograms of PIT values in the Gamma distributed example.

results. The Uniform and Uniform normal semi-stochastic methods provide worse results, in spite of having been expected to perform almost as well as the bootstrap methods in the P-P plot test. Regarding the Poisson distribution, the conclusion is similar to the outcomes of the PIT analysis: it provides too little variance to effectively predict any other distribution, and predicting it with a different method, the variance of the resulting distribution gets overly large to be considered a good fit.

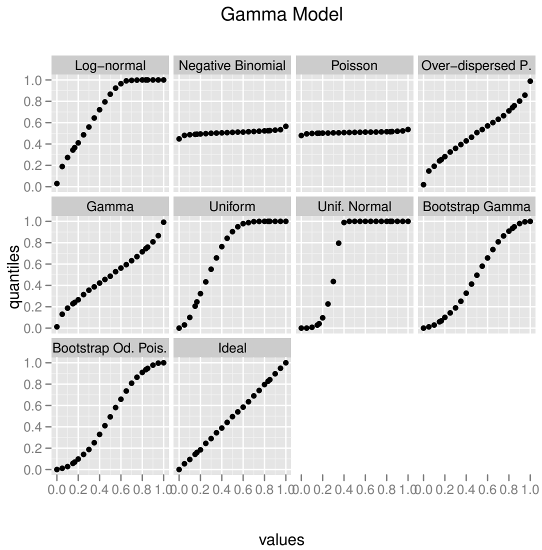


Figure 4–6: P-P Plots in the Gamma Model example.

Regarding the stochastic methods in the example, the Gamma model is clearly the best one in the score metric, followed by the overdispersed Poisson and bootstrap methods. See Figure 4–7 and 4–8, containing the boxplot representations of CRPS and ES. Table 4–9 summarises the mean score values.

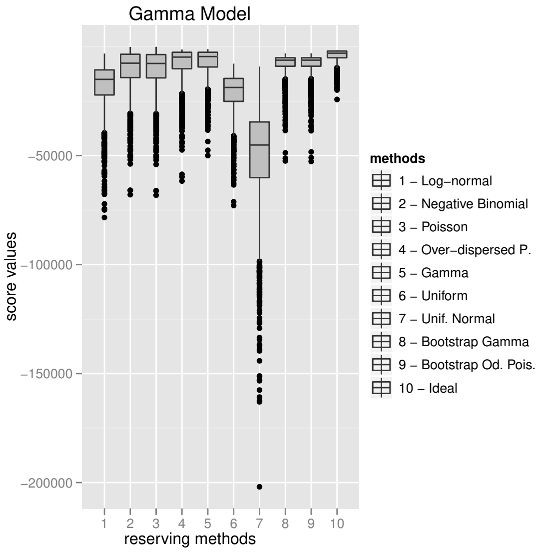


Figure 4–7: Boxplots of CRPS values in the Gamma Model example.

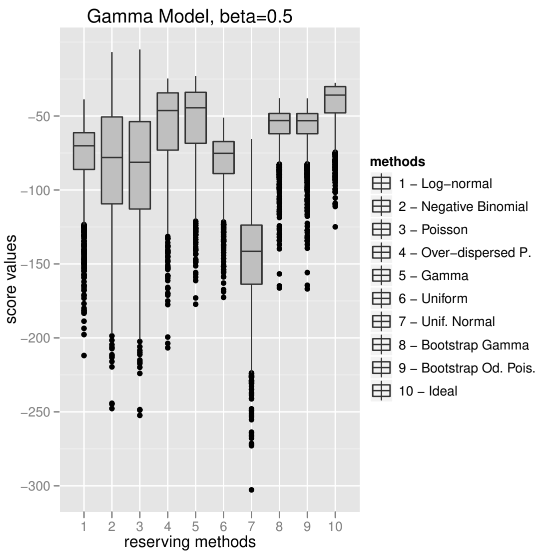


Figure 4–8: Boxplots of Energy Score values in the Gamma Model example.

Observe that results related to the Uniform and Uniform normal methods are even worse than applying any of the four non-gamma distribution models improperly. This means that despite the fact that no assumption is required about the underlying distribution when using these two evaluation techniques, results fail even against fitting inappropriate models and performing the corresponding parameter estimation. Table 4–10 highlights the inaccuracy of the applied prediction intervals. Moreover, the other run-off triangles and models have provided similar results with respect to coverage and average width. In cases where the coverage can be considered to be acceptable, the sharpness is mostly inappropriate, i.e. intervals are wide. The coverage values in conjunction with the log-normal model are closer to the 66% and 90% values, compared to the Gamma model, however, average width results are higher.

Reserving method	Coverage (%)		Average width	
	66.67% interval	90% interval	66.67% interval	90% interval
Log-normal	63.4	81.1	89035	245659
Negative Binomial	3.3	5.4	919	1562
Poisson	1.4	2.8	444	755
Overdispersed P.	48.7	71.0	15497	26322
Gamma	50.4	73.6	15506	26533
Uniform	75.4	97.0	111512	422161
Unif. Normal	95.9	100.0	287986	489479
Bootstrap Gamma	86.6	98.4	39946	70131
Bootstrap Od. Pois.	86.6	98.5	39951	70112
Ideal	66.8	89.4	13967	23821

Table 4–10: Coverage and Average width with regards to the Gamma Model.

4.6.4 Public payments data

By obtaining a detailed dataset[5] from an insurance institution, a different method has become possible. This approach deviates from the analyses on the basis of pure run-off triangles, because the data series contain the policy-based claims in contrast to the aggregate triangles. In the present case, each of the 2000 real run-off triangles were generated the following way. By drawing a 43,081-element sample with replacement from the set of 43,081 accident records, one pseudo-run-off triangle has been drawn. Besides the upper triangle part, the lower triangle of a 6×6 quadrangle accident history has become available (table C–17), i.e. the real (pseudo) ultimate claim. This step has been calculated 2000 times for each model described in the previous sections, and from that point, the simulation process was carried out similarly to the simple run-off version, as in the previous subsection. However, note that underlying scenarios differ from the other cases due to the sampling method included with replacement, which is feasible when having a set of individual claim data.

Observe that based on the PIT values in Figure 4–9, none of the methods based on distribution models can be recommended. Best histograms are yielded by the bootstrap estimation methods. Despite the fact that the uniform and the uniform

[5]Data can be downloaded from `http://amiklos.web.elte.hu/stochreserve/stochreserve.html`.

normal methods have yielded slightly worse results, these methods underperform in the cases of extreme claim values. Table 4–11 also confirms this observation, i.e. the proportion of prediction intervals of bootstrap methods are close to the 66% and 90% values. As a trade off, average width values are higher.

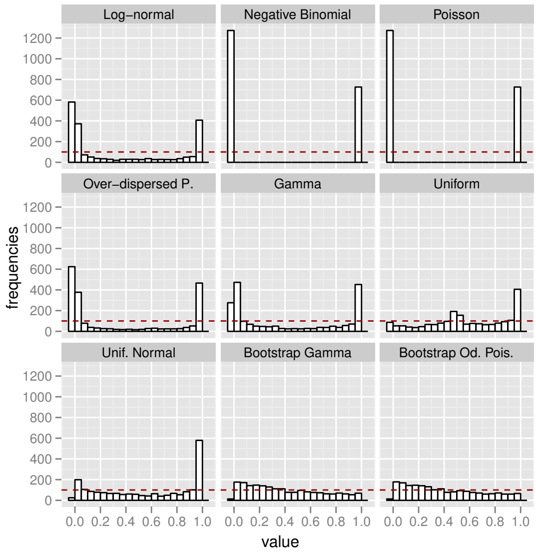


Figure 4–9: Histograms of PIT values in the Public Data case.

Reserving method	Coverage (%)		Average width	
	66.67% interval	90% interval	66.67% interval	90% interval
Log-normal	19.2	31.9	1261338486	2170190302
Negative Binomial	0.0	0.0	573564	974878
Poisson	0.0	0.0	163897	278579
Overdispersed P.	15.2	26.7	954272655	1621537701
Gamma	23.8	40.0	1395724439	2375100112
Uniform	56.2	72.5	3228640816	4738399784
Unif. Normal	38.8	59.9	1963981053	3337926040
Bootstrap Gamma	61.5	87.1	4490491606	8679547627
Bootstrap Od. Pois.	61.6	87.1	4492743160	8679552105

Table 4–11: Coverage and Average width in the Public Data case.

CRPS scoring approach has resulted in the highest figure with respect to the uniform method, see Table 4–12, despite the fact that it has not performed acceptably for other run-off triangles. Observe that energy scores are slightly better for the bootstrap methods. Furthermore, the MSEP values are the highest in the bootstrap cases, in other words, this measure implies a different ranking.

Reserving Method	CRPS (mean)	En. Sc. (mean)	MSEP (mean)	MSEP (median)
Log-normal	-1.526e+09	-27780	542.05	369.96
Negative Binomial	-1.817e+09	-38850	534.89	367.12
Poisson	-1.817e+09	-38970	534.89	367.12
Overdispersed P.	-1.579e+09	-28820	535.03	367.11
Gamma	-1.397e+09	-25880	480.34	323.97
Uniform	-1.227e+09	-23530	410.11	272.68
Unif. Normal	-1.231e+09	-23600	410.01	272.40
Bootstrap Gamma	-1.347e+09	-23470	7204.46	530.01
Bootstrap Od. Pois.	-1.347e+09	-23470	4082.06	530.80

Table 4–12: Scores and Mean Square Errors in the Public Data case (MSEP in units of 10e+16).

4.7 Conclusion

The aim of the chapter has been as follows:

1. To approach the answers to the three questions proposed in section 4.6.3.
2. To examine the score metrics applied within a claims reserving environment along with PIT, coverage, average width and msep.
3. To propose an algorithmic way of stochastic reserving model evaluation through history simulation. This is due to the lack of available real historic claims data (only a few triangles for parameterisation).
4. To present the algorithm on one real data set.

The indispensability of simulation methods has been demonstrated and proven in the comparison analysis of stochastic claims reserving models. As described, a multistage approach can be used in order to draw random scenarios and set up rankings among models on the basis of different goodness-of-fit measures. The analysed parameters and datasets imply that bootstrap gamma and bootstrap overdispersed Poisson yield the best results, and these methods are the least sensitive to the underlying distribution. It has also become clear that from the perspective of each evaluation the Uniform and Uniform normal models are considerably worse than the bootstraps. When the underlying distribution is anticipated correctly, the five parameter approximating MC methods give fair results. However, as the P-P plots and PIT values highlight, these approaches can be significantly below the quality of bootstrap techniques when being applied to the wrong distribution. Observe that rankings among the improperly chosen approximations have been calculated.

The PIT and the P-P plots suggest that the use of the Poisson model for the parameters stemming from RAA is not ideal. Having studied other datasets, the limitations of the Poisson model are rather obvious, however, from a goodness-of-fit perspective, the log-normal model was as good as the other distributions. In general, scores reasonably reflect the fit to the background distribution. Methods taken as examples in the present chapter along with the models applied within the insurance industry are not reliable with respect to prediction intervals. This negative result is less surprising due to the limited information stemming from a regular run-off triangle. In order to improve the predictions, [78] proposed the usage of Bayesian methods. Besides the methods based on aggregated claims for accident and development years, probabilistic forecasts can be done using individual contract and claim data, which is a proportionally less researched perspective. See [1, 84].

A dataset consisting of individual claims provided by an insurance company has also been utilised to test methods and scores. Note that this dataset has been included in order to propose and demonstrate a new technique to compare stochastic reserving methods. Although the mathematically sufficient establishment of the technique can be improved in the future, it might have practical relevance. One can simulate many scenarios from the past, and choose a model which best fits the claims data of the insurance institution.

In stochastic claims reserving, both in theoretical and in applied situations, it is worthwhile to test the quality of the different methods in as many ways as possible. The goal was not to construct the best stochastic claims reserving technique, but to propose an adequate methodology for comparisons in the future.

5 Stochastic Claims Reserving and Real Data

5.1 Introduction

In Chapter 4 we have seen the comparison of stochastic models based on their forecasting performance. The comparison has been done without available claim observations, except for a few triangles which were used to calibrate the random claim scenario generation that has been governed by several distributional models. In addition, the detailed observation set of one corporation has been analysed. In the present chapter the comparisons will be done based on real scenarios, using the database published in [79]. Paid and incurred claims data originate from the National Association of Insurance Commissioners (NAIC), and contain tables for six different lines of business, encompassing (1) commercial auto and truck liability and medical, (2) medical malpractice, (3) private passenger auto liability and medical, (4) product liability, (5) workers' compensation and (6) other liability. Lines of business are homogeneous groups of policies with identical coverage. Data are segmented into these clusters in order to avoid the amalgamation of claim payment run-offs with significantly different characteristics. [64] evaluates backtesting on the referred data with respect to the application of bootstrap overdispersed Poisson model.

Simulations have been performed in `R`, using packages `ChainLadder` and `rjags` for the MCMC simulations. Besides the self-written program codes, scripts published in association with [77] have been embedded into the calculations.

To our knowledge, neither the credibility bootstrap method in subsection 5.3.3, nor the collective semi-stochastic model in subsection 5.3.5 has been discussed in peer-reviewed journals (before [72]). Two of the models incorporate experience ratemaking from the claims history of an entire community of companies. One step further is harnessing collective data in order to improve individual (insurance company level) prediction reliabilities, requiring the coordination of regulatory authorities as data collectors and distributors.

To summarise the novelties communicated by the present chapter: (1) Metrics in actuarial reserving such as CRPS, coverage and sharpness of several models to analyse their performance and determine an order of appropriateness have been presented by [4] on simulated data. Here we apply all the calculations on *actual triangles* from multiple risk groups. (2) PIT has already been applied by [77] on stochastic models, here we continue presenting the calculations involving further methods not covered elsewhere (credibility bootstrap, bootstrap Munich, semi-stochastic). (3) Two new models are introduced, credibility bootstrap in 5.3.3 and collective semi-stochastic in 5.3.5. (4) We emphasise the importance of an algorithmic way of model selection from competing peers in section 5.4. (5) Models based on individual company information only (single triangle) are also compared with collective ones (multiple triangles and credibility). The present chapter also sheds light on the potential of oversight data collection and possible application on multiple triangles. (6) Scripts published by [77] are developed further with new code chunks and made available online.

The chapter is structured as follows: Section 5.2 contains the expository description of insurance data published by the NAIC and used for comparative analysis, consisting of observations of claims and premiums from hundreds of insurance institutions. Section 5.3 consists of diverse reserving models, not discussed in Chapter 4, and a number of which are applied widely in the insurance industry. Having approached the original, claims reserving problem as a probabilistic forecast, section 5.4 provides insight into the calculated values of the five measures introduced in section 4.5. The section includes the validation of individual models from the angle of the five indicators. Section 5.5 concludes the chapter. The present chapter is founded on [72].

5.2 Data

Open source data enables the validation of methods on real loss figures. The National Association of Insurance Commissioners (NAIC) published data tables consisting of the names of insurance institutions, incurred and paid loss per accident year and per development year, and earned premiums per contract year. [79] published these tables along with the article.

Historical values applied in the present chapter concern the run-off triangles built up by paid and incurred losses. Six different lines of business can be distinguished; (1) commercial auto and truck liability and medical, (2) medical malpractice, (3) private passenger auto liability and medical, (4) product liability, (5) workers' compensation and (6) other liability, with a variable number of corporations contributing to the dataset. Business lines correspond to homogeneous segments of insurance portfolios, which are addressed separately for the reason that they generally show distinct run-off behaviour. Hence, clusters on the basis of coverage type are made in order not to

	business line	*number of observations*
(1)	commercial auto and truck liability and medical	158
(2)	medical malpractice	34
(3)	private passenger auto liability and medical	146
(4)	product liability	70
(5)	workers' compensation	132
(6)	other liability	239

Table 5–1: Number of observations (insurance institutions) in the datasets.

amalgamate different run-off characteristics. Let one observation mean the loss triangle associated with one insurance company, see Table 5–1.

In fact, accident years cover a 10-year time span between 1988 and 1997, with a 10-year development lag for each accident year. In other words, not only the triangle values above (and including) the anti-diagonal are available (Table 5–2), but the entire rectangle in each case. From a validation perspective, it is crucial that the actual ultimate claim values, i.e. the lower triangles are known (Table 5–3).

	1	2	3	4	5	6	7	8	9	10
1988	5407	14422	19063	22447	24142	25404	26829	27202	27443	27449
1989	6279	15031	21203	25697	27807	28726	29173	29375	29444	
1990	7256	15923	20701	24963	27847	29274	30163	30656		
1991	5028	10345	15042	18837	21708	22808	23465			
1992	5712	11809	18198	22000	26306	27168				
1993	7413	16798	24570	30420	33803					
1994	10868	23205	31171	39702						
1995	10143	24336	32406					?		
1996	9596	21831								
1997	9076									

Table 5–2: Cumulative paid loss triangle observed in the past (commercial auto dataset, group code 2712).

	1	2	3	4	5	6	7	8	9	10
1988										
1989										29459
1990									30691	30749
1991								24243	25020	25061
1992							27525	27888	27951	28042
1993						34881	35984	36313	36509	36524
1994					43225	45450	46662	47034	47027	47186
1995				38533	42552	44730	45197	45362	45516	45765
1996			27594	31228	33710	36683	36417	37068	37086	37141
1997		17689	23270	29846	33532	35205	35410	35443	35501	35540

Table 5–3: Cumulative paid loss triangle observed in the future (commercial auto dataset, group code 2712).

Ultimate claim values range from zero to millions in extreme cases, see Table 5–4 for paid losses, implying magnitudinal diversity in the set of companies in terms of reserves. In fact, only a few outliers can be found with *negative* total claims, which we consider the less reliable part of the data set. These instances have been taken out of the analysis. Therefore, a natural and far not trivial question is whether or not to apply a normalisation on the run-off triangles, in order to make reserving models reasonably comparable with each other by mitigating the heterogeneity of the underlying figures. For instance, this can be achieved by multiplying each triangle by different constants to make ultimate reserves equal to a unit value. Several pitfalls come with the scaling: applying a discrete model such as the overdispersed Poisson model (family) on triangles consisting of small numbers, the estimation will be useless if the Poisson parameter is close enough to zero to make future claim increments equal to zero with high probability. As a matter of fact, this issue can be remediated by choosing an appropriately large normalising constant. The standardisation of such overdispersed Poisson data has been extensively discussed in the past in connection with stochastic reserving. Each of the run-off triangle elements are normalised by a volume measure related to the accident year, i.e. each incremental or cumulative claim in row i is divided by a weight $w_i > 0$. This exposure volume can be the number of reported claims in accident year i, see [110]. Another convention is to choose the earned premium volume or the number of policies, see [96].

The second and more contradictory argument against scaling is embedded in the data: large companies likely provide more robust claim records than their smaller counterparts, i.e. it is rational to take them into account with larger weights, which is ensured by the larger reserve values. Hence, the question is whether to allow institutions to contribute to the total loss values according to their reserve volumes, or compose a *democratic* aggregate observation set with a similar contribution from each institution in terms of ultimate claim. An intermediate solution can be a nonconstant rescaling of data. In loss reserving calculations, the author in [95] applies normalisation in order to mitigate the heterogeneity of the data. Present calculations leave the original figures as they are.

5.3 Claims reserving models revisited

The (1) bootstrap models with Gamma and overdispersed Poisson background, which have been discussed in chapter 4, will be applied in the present one as well. In this section four new, conceptually distinct modelling approaches are explained in claims reserving, where in some of the cases, the model refers to a method family

	Min.	Median	Mean	Max.
commercial auto and truck liability and medical	-1	3 906	50 820	2 227 000
medical malpractice	0	15 600	95 370	883 900
private passenger auto liability and medical	0	19 810	818 100	91 360 000
product liability	0	316	19 430	750 300
workers' compensation	0	8 828	101 900	1 837 000
other liability	-115	913	20 460	2 191 000

Table 5–4: Ranges of paid losses per business line.

rather than a single one. These are the (2) Bayesian models using MCMC techniques, (3) credibility models, including a newly introduced one combined with bootstrapping, (4) original Munich Chain Ladder and its bootstrapped modification and (5) a modified semi-stochastic model.

Notations the reader frequently encounters in this section are the following: I and J denote the number of occurrence and development years in the triangles (and quadrangles), i.e. they stand for the dimensions. Let C^I and C^P denote the incurred and paid triangles in subsection 5.3.4. Avoid confusing the superscript in C^I, which stands for 'Incurred', with the I number of rows in the triangle. If the paid or incurred indicatives are not relevant from a technical perspective, they will not be marked. Superscript (k) in connection with cumulative triangle element $C_{i,j}$ means that the value is related to company k. $\mathcal{D}_k$ stands for the upper run-off triangle of the kth company, i.e. the claims data acquired until the time of reserve calculation. The latter notation is used in subsection 5.3.3 in order to avoid confusion with other indices.

5.3.1 Bootstrap models

See section 4.4.2. In the present chapter, Step B is simualted 1000 times.

5.3.2 Bayesian models using MCMC

Two methods based on Markov Chain Monte Carlo simulation that follow a Bayesian concept are presented by [77]. The author made the self-prepared R codes public in order to facilitate the replication of results. These code chunks have been embedded into the set of codes supporting the analysis. Models with MCMC sampling are the most computation-intensive ones among the modelling principles.

5.3.2.1 Correlated chain ladder model

In the correlated chain ladder (CCL) model incurred claims are the basis of calculation, in the form of cumulative losses. The motivation is to address the possible underestimation of ultimate claim variability in the original Mack model [68]. The underlying assumption is that the unknown losses $\tilde{C}_{i,j}$ are governed by the log-normal distribution. See [77] for the detailed model assumptions. Let

1. $\alpha_i \sim N\left(\log(\text{Premium}_i) + \text{logelr}, \sqrt{I}\right)$ with logelr $\sim Unif(-1, 0.5)$ (or precision)
2. $\beta_j \sim Unif(-5, 5)$ $1 \leq j < I$
3. $\varrho \sim Unif(-1, 1)$
4. $\mu_{1,j} = \alpha_1 + \beta_j$ and $\mu_{i,j} = \alpha_i + \beta_j + \varrho\left(\log(C_{i-1,j}) - \mu_{i-1,j}\right)$ for $1 < i \leq I$ and $1 \leq j \leq I$ ($\beta_I = 0$ to prevent overparameterisation)
5. $\tilde{C}_{i,j} \sim LN\left(\mu_{i,j}, \sigma_j\right)$, assuming strictly monotonically decreasing variances $\sigma_1 > \sigma_2 > \cdots > \sigma_I$.
6. $\sigma_j = \sum\limits_{l=j}^{I} a_l \; \forall j$, where $a_l \sim Unif(0, 1)$

If the parameters above are given, ϱ represents the correlation between the two subsequent cumulative claim values pertaining to one occurrence year, i.e. $Corr\left(\log \tilde{C}_{i,j-1}, \log \tilde{C}_{i,j}\right)$. The predictive distribution for the ultimate claim $\sum\limits_{i=1}^{I} C_{i,J}$ is a result of the MCMC simulation. In the first step, create a $\alpha_i, \beta_j, \varrho$ sample $(\forall i, \forall j)$. In the second step, generate $C_{i,J}$ values $i = (1), 2, \ldots, I$ by setting $\mu_{i,J}$ step-by-step consecutively from $C_{i-1,J}$. Repeat the process in order to achieve a sufficiently large number of sample (10,000 in the concrete examples).

5.3.2.2 Correlated incremental trend model

The second model is built on the incremental paid loss amounts rather than the incurred claims, and has a distribution skewed to the right. For skew-normal distribution see [42].

Definition 5.1 (skew-normal distribution) Let $\mu \in \mathbb{R}$ be a location, $\omega \in (0, \infty)$ a scale and $\delta \in (-1, 1)$ a shape parameter. Let Z be a random variable of Truncated $\text{Normal}_{[0,\infty)}(0, 1)$ and ε of standard normal variable, independent from each other. Then $\xi \sim SN(\alpha)$ is said to be a skew-normal distributed variable with skewness parameter $\alpha = \frac{\delta}{\sqrt{1-\delta^2}}$, if it is the $\xi \stackrel{d}{=} \delta Z + \sqrt{1 - \delta^2}\varepsilon$ mixture of the previous variables. Applying an affine transformation on the just defined variable in order to adjust it to arbitrary location and scale parameters yields the skew-normal distribution in the general sense; $Y \sim SN(\mu, \omega^2, \alpha)$, if $Y = \mu + \omega\delta Z + \omega\sqrt{1 - \delta^2}\varepsilon$.

[77] points to the issue that a skew-normal distribution has the skewness of a truncated normal variable in the extreme case, which still may not reflect the real skewness stemming from the loss data, creating the demand for an even more skewed distribution to be applied instead of the truncated normal.

Definition 5.2 (mixed lognormal-normal) Let Z be a log-normal random variable with scale and shape parameters μ and σ, and let the mixture $\delta Z + \sqrt{1 - \delta^2}\varepsilon$ be called mixed lognormal-normal distribution with parameters μ, δ, σ.

Hence, in the second, correlated incremental trend (CIT) model, let the following objects be introduced.

1. $\mu_{i,j} = \alpha_i + \beta_j + \tau \cdot (i + j - 1)$, $1 \leq i \leq I$ and $1 \leq j \leq I$
2. $Z_{i,j} \sim LN(\mu_{i,j}, \sigma_j)$ with strictly increasing σ values in j, $\sigma_1 < \sigma_2 < \ldots < \sigma_I$
3. $\tilde{X}_{1,j} \sim N(Z_{1,j}, \delta)$ and $\tilde{X}_{i,j} \sim N\left(Z_{i,j} + \varrho \cdot (\tilde{X}_{i-1,j} - Z_{i-1,j}) \cdot e^{\tau}, \delta\right)$ $i > 1$, where τ stands for the trend factor on payments

In the calculations on real data, prior distributions are assigned as follows.

1. $\alpha_i \sim N(\log \text{Premium}_i + \text{logelr}, \sqrt{I})$, where logelr $\sim Unif(-5, 1)$
2. $\beta_j \sim Unif(0, I)$ for $j \in \{1, \ldots, 4\}$, $\beta_j \sim Unif(0, \beta_{j-1})$ for $j \in \{5, \ldots, I-1\}$ and $\beta_I = 0$
3. $\varrho \sim Unif(-1, 1)$
4. $\tau \sim N\left(0, \frac{1}{1000}\right)$
5. $\sigma_1^2 \sim Unif(0, \frac{1}{2})$, $\sigma_j^2 \sim Unif(\sigma_{j-1}^2, \sigma_{j-1}^2 + 0.1)$ for $j > 1$
6. $\delta \sim Unif(0, \text{Average Premium})$

Another model in the referred monograph, called changing settlement rate model, may address the phenomenon of accelerating claim settlements, driven by technological changes.

5.3.3 Credibility models

The present subsection contains the basic idea of credibility theory and its connection with claims reserving. By combining this idea with bootstrapping, a new reserving model is introduced.

Papers [15, 16] contain the original concept of experience ratemaking. The core principle is to make use of the available information from sources outside of the sample, but somehow related to it, and combine the two datasets in order to get a more reliable approximation of unknown characteristics. Considering one business line, in order to create the claim forecast of one particular triangle, the other run-off triangles of the same group are also taken into account. From another angle, the model consists of 2 urns, where we pick the risk parameter ϑ from the first one, which determines the value sampled from the second urn. [97] proposes credibility based stochastic reserving driven by the idea that data from peer counterparty insurers can lead to an improvement of prediction reliability.

To the analogy of the Mack Chain Ladder methodology [68], construct the following model assumptions, applying Bayesian thinking.

Assumption 5.1 (credibility)

(C 1) Let each unknown chain ladder factor be a positive random variable F_j for $\forall j \in \{1, \dots, J-1\}$, F_i independent of F_j for $\forall i \neq j$.
(C 2) $C_{0,j}, \dots, C_{I,j}$ are conditionally independent of $\underline{F}$.
(C 3) The conditional distribution of $\frac{C_{i,j+1}}{C_{i,j}}$ under the constraint $\sigma\left(\{F_0, \dots, F_j, C_{i,0}, \dots, C_{i,j}\}\right)$ depends only on $\sigma\left(\{F_j, C_{i,j}\}\right)$. Furthermore, conditional expectation and variance are

$$E\left(\frac{C_{i,j+1}}{C_{i,j}}|F_j, C_{i,j}\right) = F_j \quad \text{and} \quad Var\left(\frac{C_{i,j+1}}{C_{i,j}}|F_j, C_{i,j}\right) = \frac{\sigma_j^2(F_j)}{C_{i,j}}.$$

Recall from Bayesian statistics that for an arbitrary random variable ξ and array of observations $\underline{X}$, the linear Bayesian estimator satisfies $\underset{\hat{\xi}:\hat{\xi}=\sum_i a_i X_i + const}{\arg\min} E\left((\hat{\xi}-\xi)^2|\underline{X}\right)$.
Also recall from [44] definition 5.3 and theorem 5.1.

Definition 5.3 (credibility based predictor) The credibility based predictor of the ultimate claim $C_{i,J}$ given $\mathcal{D}$ is

$$C_{i,J}^{cred} = C_{i,I-i} \prod_{j=I-i}^{J-1} F_j^{cred},$$

where

$$F_j^{cred} = \underset{\hat{F}_j:\hat{F}_j=\sum_{i=1}^{I-j} a_{i,j} Y_{i,j} + const}{\arg\min} E\left((\hat{F}_j - F_j)^2|\mathcal{B}(j)\right)$$

and $Y_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$, $\mathcal{B}(j) = \{C_{i,k} : i+k \leq I+1, k \leq j\} \subset \mathcal{D}$ the subset of upper triangle information.

Given the multiplicative structure of the ultimate claim estimator it may not be appropriate to call it simply a credibility estimator, which is by definition a linear function of the observations, hence the credibility based appellation.

Theorem 5.1 *The credibility estimators of the development factors are given by*

$$F_j^{cred} = \alpha_j \hat{F}_j + (1-\alpha_j) f_j,$$

where $\hat{F}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}$, $f_j = E(F_j)$, $\alpha_j = \frac{\sum_{i=1}^{I-j} C_{i,j}}{\sum_{i=1}^{I-j} C_{i,j} + \frac{\sigma_j^2}{\tau_j^2}}$, $\sigma_j^2 = E(\sigma_j^2(F_j))$ *and* $\tau_j^2 = Var(F_j)$. *The latter two are the structural parameters (or credibility factors and their quotient,* $\kappa_j = \frac{\sigma_j^2}{\tau_j^2}$ *is the credibility coefficient).*

For the mean square error of prediction it is also true that $msep(F_j^{cred}) = (1 - \alpha_j)\tau_j$, *see definition 4.10.*

Proof: See [44].

Data concerning the credibility factor in particular are not available in general. In the present work these parameters are approximated on the basis of claim triangles published by several companies.

From regulatory perspective it is extremely important to understand how the inflowing data can be exploited in order to support insurance institutions with reliable information. Financial regulatory authorities tend to collect an increasing amount of detailed data for the purpose of gaining insight into insurance institutions' solvency. In Europe, the European Insurance and Occupational Pensions Authority (EIOPA) provides local regulators with guidance and collects statistical and financial data from several countries. Besides transparency, the information enables the adequate support of corporations by providing them with processed data to their benefit. This is where credibility models have an untapped potential. The question whether or not to use collective experience to improve individual approximations is particularly relevant due to the fact that regulatory authorities collect vast amount of information from insurance companies. The aggregated data might be of value to share with the contributors, enabling more precise solvency evaluations.

Let $C_{i,j}^{(k)}$ stand for the cumulative payment or incurred claim value with occurrence year i and development year j with respect to company k. In general, for simplicity's sake it is supposed that for each insurance institution the triangle dimensions are equal, moreover, $I = I^{(1)} = I^{(2)} = \ldots = I^{(n)}$. n denotes the number of companies observed in a homogeneous risk group and $I^{(k)}$ the dimension of the kth triangle.

The parameter estimation of credibility factors is constructed in accordance with section 4.8 in [17]. Let index j be fixed and let $S_j^{(k)}$ $k \in \{1, \ldots, n\}$ be defined for each triangle as

$$S_j^{(k)} = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j}^{(k)} \left(\frac{C_{i,j+1}^{(k)}}{C_{i,j}^{(k)}} - \frac{\sum\limits_{r=1}^{I-j} C_{r,j+1}^{(k)}}{\sum\limits_{r=1}^{I-j} C_{r,j}^{(k)}} \right)^2 . \tag{5.1}$$

Observe that

$$\begin{aligned} S_j^{(k)} &= \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j}^{(k)} \left(\frac{C_{i,j+1}^{(k)}}{C_{i,j}^{(k)}} - F_j + F_j - \frac{\sum\limits_{r=1}^{I-j} C_{r,j+1}^{(k)}}{\sum\limits_{r=1}^{I-j} C_{r,j}^{(k)}} \right)^2 = \\ &= \frac{1}{I-j-1} \sum_{i=1}^{I-j} \left(C_{i,j}^{(k)} \left(\frac{C_{i,j+1}^{(k)}}{C_{i,j}^{(k)}} - F_j \right)^2 - \sum_{r=1}^{I-j} C_{r,j}^{(k)} \left(\frac{\sum\limits_{r=1}^{I-j} C_{r,j+1}^{(k)}}{\sum\limits_{r=1}^{I-j} C_{r,j}^{(k)}} - F_j \right)^2 \right), \end{aligned} \tag{5.2}$$

which implies that $E(S_j^{(k)}|F_j) = \sigma_j^2(F_j)$ is in line with assumption 5.1. Hence, $E(S_j^{(k)}) = E(E(S_j^{(k)}|F_j)) = E(\sigma_j^2(F_j)) = \sigma_j^2$ for each j, i.e. $S_j^{(k)}$ provides an unbiased estimator for σ_j^2. Taking the average of $S_j^{(k)}$ values for all the companies results in an unbiased estimator of σ_j^2:

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j}^{(k)} \left(\frac{C_{i,j+1}^{(k)}}{C_{i,j}^{(k)}} - \frac{\sum\limits_{l=1}^{I-j} C_{l,j+1}^{(k)}}{\sum\limits_{l=1}^{I-j} C_{l,j}^{(k)}} \right)^2 . \tag{5.3}$$

It can also be shown with further calculations that $\hat{\tau}_j^2$ is an unbiased estimator of τ_j^2:

$$\hat{\tau}_j^2 = c_j \left(\frac{n}{n-1} \sum_{k=1}^{n} \frac{\sum\limits_{i=1}^{I-j} C_{i,j}^{(k)}}{\sum\limits_{l=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j}^{(l)}} \left(\frac{\sum\limits_{i=1}^{I-j} C_{i,j+1}^{(k)}}{\sum\limits_{i=1}^{I-j} C_{i,j}^{(k)}} - \frac{\sum\limits_{l=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j+1}^{(l)}}{\sum\limits_{l=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j}^{(l)}} \right)^2 - \frac{n \cdot \hat{\sigma}_j^2}{\sum\limits_{k=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j}^{(k)}} \right) \tag{5.4}$$

with $c_j = \frac{n-1}{n} \left(\sum\limits_{k=1}^{n} \frac{\sum\limits_{i=1}^{I-j} C_{i,j}^{(k)}}{\sum\limits_{l=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j}^{(l)}} \cdot \left(1 - \frac{\sum\limits_{i=1}^{I-j} C_{i,j}^{(k)}}{\sum\limits_{l=1}^{n} \sum\limits_{i=1}^{I-j} C_{i,j}^{(l)}} \right) \right)^{-1}$.

Parameter τ_j needs extra attention, having observed that the estimator below can attain negative values, not only in an extremely theoretical sense, but in real world trajectories, as well. For that reason, let the approximation be capped by 0 from below.

$$\hat{\tau}_j^2 = \max\left(0, \hat{\hat{\tau}}_j^2\right). \tag{5.5}$$

Furthermore, let the estimator of $f_j = E(F_j)$ be

$$\hat{f}_j = \sum_{k=1}^{n} \frac{\alpha_j^{(k)}}{\sum_{l=1}^{n} \alpha_j^{(l)}} \cdot \frac{\sum_{i=1}^{I-j-1} C_{i,j+1}^{(k)}}{\sum_{i=1}^{I-j-1} C_{i,j}^{(k)}} \tag{5.6}$$

Method 5.1 (Credibility Bootstrap)

(Step 1) Take the pool of $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ run-off triangle observations. Estimate $f_j, \sigma_j^2, \tau_j^2$ for each j according to equations 5.6, 5.3, 5.5.

(Step 2) With respect to each company, exchange the chain ladder factors with the credibility chain ladder factors.

(Step 3) Apply the bootstrap overdispersed Poisson model with the credibility chain ladder factors. (subsection 4.4.2).

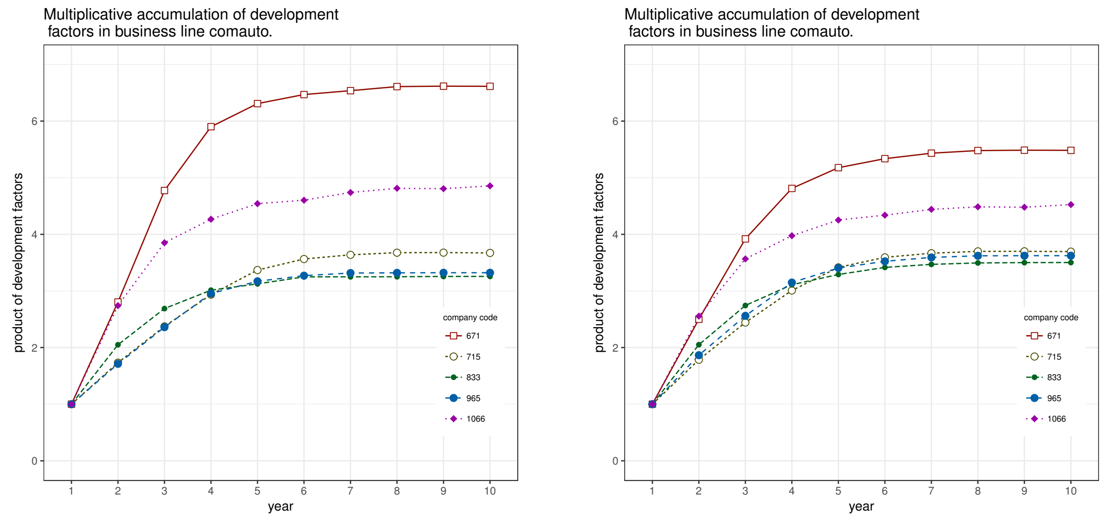
As an illustration of the outcome of the first two steps in the credibility bootstrap method, consider a few arbitrarily selected companies in one business line. The cumulative product of the λ_i development factors can be seen in Figure 5–1 (a) for each, in the sense that the function value of the first year is equal to 1, and $1 \cdot \lambda_1 \cdot \ldots \cdot \lambda_{k-1}$ for year k $(k = 2, \ldots, 10)$.

Figure 5–1 (b) presents the same institutions as the previous figure, but with credibility adjustment, i.e. instead of the original λ_i values, the F_j^{cred} developments in a similarly product-based pattern. Observe the narrowing range of individual patterns.

5.3.4 Munich chain ladder model

5.3.4.1 Original model

Observe that all the reserving models enumerated so far are operated with one run-off triangle, be it either the paid or the incurred one. The question naturally arises why not to use both triangles at the same time, doubling the volume of the information and, hopefully upgrading the quality of prediction. [87] introduced the Munich Chain Ladder (MCL) algorithm, which takes both paid and incurred cumulative data into account, assuming correlation between paid and incurred claims that stem from different accident years.



(a) (b)

Figure 5–1: Multiplicative accumulation of development factors (a) *without* and (b) *with* credibility adjustment.

Notation 5.1

1. Let $F^P_{i,j} = \frac{C^P_{i,j+1}}{C^P_{i,j}}$ stand for the regular chain ladder development factors in the paid, and $F^I_{i,j} = \frac{C^I_{i,j+1}}{C^I_{i,j}}$ in the incurred triangle, where $i = 1, \ldots, I$ and $j = 1, \ldots, J - i$. Let $Q_{i,j} = \frac{C^P_{i,j}}{C^I_{i,j}}$ and $Q^{-1}_{i,j} = \frac{C^I_{i,j}}{C^P_{i,j}}$, $i = 1, \ldots, I$ and $j = 1, \ldots, J - i + 1$, be the ratios of paid and incurred claims, or (P/I) and (I/P) ratios.
2. Let generated σ-fields $\mathcal{P}_i(k) = \sigma\{C^P_{i,j} : j \leq k\}$ and $\mathcal{I}_i(k) = \sigma\{C^I_{i,j} : j \leq k\}$ be the information acquired until development year k, related to claims in accident year i. Let $\mathcal{B}_i(k)$ denote the combined knowledge $\sigma\{C^P_{i,j}, C^I_{i,j} : j \leq k\}$.

Assumption 5.2 (Munich chain ladder)

(A) (Expectations) There exist positive development factors f^P_j and f^I_j such that

$$E(F^P_{i,j}|\mathcal{P}_i(j)) = f^P_j \text{ and } E(F^I_{i,j}|\mathcal{I}_i(j)) = f^I_j$$

$\forall i \in \{1, \ldots, I\}$ and $\forall j \in \{1, \ldots, J\}$. Furthermore, there exist q_j and q^{-1}_j such that

$$E(Q_{i,j}|\mathcal{I}_i(j)) = q_j \text{ and } E(Q^{-1}_{i,j}|\mathcal{P}_i(j)) = q^{-1}_j.$$

(B) (Variances) There exist non-negative constants σ^P_j and σ^I_j such that

$$Var(F^P_{i,j}|\mathcal{P}_i(j)) = \frac{(\sigma^P_j)^2}{C^P_{i,j}} \text{ and } Var(F^I_{i,j}|\mathcal{I}_i(j)) = \frac{(\sigma^I_j)^2}{C^I_{i,j}}$$

$\forall i \in \{1, \ldots, I\}$ and $\forall j \in \{1, \ldots, J\}$. Furthermore, there exist ϱ_j^I and ϱ_j^P such that

$$Var(Q_{i,j}|\mathcal{I}_i(j)) = \frac{(\varrho_j^I)^2}{C_{i,j}^I} \text{ and } Var(Q_{i,j}^{-1}|\mathcal{P}_i(j)) = \frac{(\varrho_j^P)^2}{C_{i,j}^P}.$$

(C) (Independence) Occurrence years are independent, i.e. sets

$$\{C_{1,j}^P, C_{1,j}^I : j = 1, \ldots, J\}, \ldots, \{C_{I,j}^P, C_{I,j}^I : j = 1, \ldots, J\}$$

are stochastically independent.

(D) (Correlations) Generally, let $Res(\xi|\mathcal{A}) = \frac{\xi - E(\xi|\mathcal{A})}{\sqrt{Var(\xi|\mathcal{A})}}$ denote the conditional residual of random variable ξ given σ-algebra $\mathcal{A}$. There exist λ^P and λ^I constants such that

$$E\left(Res(F_{i,j}^P|\mathcal{P}_i(j))|\mathcal{B}_i(j)\right) = \lambda^P \cdot Res(Q_{i,j}^{-1}|\mathcal{P}_i(j))$$

and

$$E\left(Res(F_{i,j}^I|\mathcal{I}_i(j))|\mathcal{B}_i(j)\right) = \lambda^I \cdot Res(Q_{i,j}|\mathcal{I}_i(j)).$$

Rearranging the equations results in forms

$$E(F_{i,j}^P|\mathcal{B}_i(j)) = f_j^P + \lambda^P \cdot \frac{\sqrt{Var(F_{i,j}^P|\mathcal{P}_i(j))}}{\sqrt{Var(Q_{i,j}^{-1}|\mathcal{P}_i(j))}} \cdot \left(Q_{i,j}^{-1} - E(Q_{i,j}^{-1}|\mathcal{P}_i(j))\right) \quad (5.7)$$

and

$$E(F_{i,j}^I|\mathcal{B}_i(j)) = f_j^I + \lambda^I \cdot \frac{\sqrt{Var(F_{i,j}^I|\mathcal{I}_i(j))}}{\sqrt{Var(Q_{i,j}|\mathcal{I}_i(j))}} \cdot \left(Q_{i,j} - E(Q_{i,j}|\mathcal{I}_i(j))\right). \quad (5.8)$$

5.3.4.2 Bootstrapping the Munich chain ladder

In its original form the MCL method fails to establish distributions for ultimate paid or incurred claim values and thus to enable the analysis of their stochastic behaviour. Recalling the application of bootstrap techniques, [66] suggests a plausible solution to generate random outcomes by drawing random samples from the four residual sets in the MCL procedure. Let $r_{ij}^P, r_{ij}^I, r_{ij}^Q, r_{ij}^{Q^{-1}}$ denote the residuals $Res(C_{i,j}^P|\mathcal{P}_i(j))$, $Res(C_{i,j}^I|\mathcal{I}_i(j))$, $Res(Q_{i,j}|\mathcal{I}_i(j))$, $Res(Q_{i,j}^{-1}|\mathcal{P}_i(j))$, where $i = 1, \ldots, I-1$, $j = 1, \ldots, I-i$ for the first two, and one anti-diagonal larger for the (P/I) and (I/P) residuals.

The idea is to correct the bootstrap bias in each element by multiplying the residuals by $\sqrt{\frac{I-j}{I-j-1}}$, then creating a new bootstrap sample by choosing random elements with replacement from the set of $\{r_{ij}^P, r_{ij}^I, r_{ij}^Q, r_{ij}^{Q^{-1}}\}$ 4-tuples. In one iteration process, values relevant in the MCL method are reversely calculated, along with the adjusted development factors $E(\tilde{F_{i,j}}^P|\mathcal{B}_i(j))$ and $E(\tilde{F_{i,j}}^I|\mathcal{B}_i(j))$. At the end of one iteration, the lower triangle is recursively completed by normal random variables:

$$C_{i,j}^P \sim N\left(\hat{\hat{\varrho}}_{i,j-1}^P \hat{C}_{i,j-1}^P, (\hat{\hat{\sigma}}_{j-1})^2 \hat{C}_{i,j-1}^P\right),$$

and analogously for incurred data

$$C_{i,j}^I \sim N\left(\hat{\hat{\varrho}}_{i,j-1}^I \hat{C}_{i,j-1}^I, (\hat{\hat{\sigma}}_{j-1})^2 \hat{C}_{i,j-1}^I\right).$$

5.3.4.3 Applicability and limitations

A practical drawback of the model which may materialise during reserve calculations is that variance parameters σ_j and ϱ_j can attain extremely low values, even zero. It means that their ratio can be a large number, which contributes to the conditional development factor, see Assumptions 5.2 (D), eventually resulting in unrealistic ultimate claims.

To give an example from the actually documented NAIC figures, see paid Table 5–5 and incurred Table 5–6 triangles from the commercial automobile insurance claims of a company.

	1	2	3	4	5	6	7	8	9	10
1988	126	256	326	369	489	489	489	489	490	490
1989	169	313	364	501	561	573	573	557	557	
1990	237	402	582	695	711	708	713	742		
1991	461	602	643	764	804	815	815			
1992	413	694	853	1204	1274	1352				
1993	802	1171	1415	1643	1823					
1994	1044	1528	1722	2002						
1995	829	1320	1579							
1996	1109	1786								
1997	1443									

Table 5–5: Cumulative paid loss triangle observed in the past (commercial auto data set, group code 8079).

Evaluating the variance parameters defined in Assumptions 5.2 (B), it becomes clear that for higher js, ϱ_j^P and ϱ_j^I gets close to zero. The unbiased parameter estimators

$$(\hat{\sigma}_j^P)^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j}^P (F_{i,j}^P - \hat{f}_j^P)^2, \quad (\hat{\varrho}_j^P)^2 = \frac{1}{I-j} \sum_{i=1}^{I-j+1} C_{i,j}^P (Q_{i,j}^{-1} - \hat{q}_j^{-1})^2 \quad (5.9)$$

	1	2	3	4	5	6	7	8	9	10
1988	351	364	347	398	489	489	489	489	490	490
1989	294	436	617	611	573	573	573	557	557	
1990	810	804	807	802	719	741	748	742		
1991	860	852	918	840	814	815	815			
1992	874	1276	1262	1400	1493	1444				
1993	2031	1860	1963	1990	2005					
1994	2293	2291	2222	2170						
1995	2027	1901	1988							
1996	2650	2833								
1997	3379									

Table 5–6: Cumulative incurred loss triangle observed in the past (commercial auto data set, group code 8079).

and

$$(\hat{\sigma}_j^I)^2 = \frac{1}{I-j-1}\sum_{i=1}^{I-j} C_{i,j}^I (F_{i,j}^I - \hat{f}_j^I)^2, \quad (\hat{\varrho}_j^I)^2 = \frac{1}{I-j}\sum_{i=1}^{I-j+1} C_{i,j}^I (Q_{i,j} - \hat{q}_j)^2 \quad (5.10)$$

may result in almost zero numbers due to the fact that as index j approaches J, each sum of the four estimators can be close or equal to zero, see Table 5–7. Thus, excessive fractions $\frac{\hat{\sigma}_j^P}{\hat{\varrho}_j^P}$ and $\frac{\hat{\sigma}_j^I}{\hat{\varrho}_j^I}$ yield degenerate MCL development factor estimations $\hat{f}_j^P + \hat{\lambda}^P \cdot \frac{\hat{\sigma}_j^P}{\hat{\varrho}_j^P}(Q_{i,j}^{-1} - \hat{q}_j^{-1})$ and $\hat{f}_j^I + \hat{\lambda}^I \cdot \frac{\hat{\sigma}_j^I}{\hat{\varrho}_j^I}(Q_{i,j} - \hat{q}_j)$, exceeding any upper bound.

	1	2	3	4	5	6	7	8	9
σ^P	3.4e+00	2.4e+00	3.5e+00	2.0e+00	7.1e-01	5.1e-01	9.8e-01	1.0e+00	9.8e-01
σ^I	5.7e+00	3.1e+00	2.6e+00	2.2e+00	7.6e-01	3.2e-01	3.8e-01	3.7e-01	5.8e-01
ϱ^P	7.8e+00	4.8e+00	4.0e+00	1.9e+00	2.2e+00	9.8e-01	6.4e-01	3.6e-15	3.6e-15
ϱ^I	2.1e+00	2.3e+00	2.5e+00	1.5e+00	1.9e+00	9.3e-01	6.2e-01	3.6e-15	3.6e-15

Table 5–7: Variance assumption parameters. (The ϱ^P, ϱ^I parameters become zero for development steps 8 and 9.)

Such estimators contribute to the approximate reserves in Table 5–8, see columns MCL paid and incurred. The astronomical values are the direct result of the parameter calculation according to the closed formulas in Equation 5.9 and 5.10. Hence, as an alternative, change $\frac{\hat{\sigma}_j^P}{\hat{\varrho}_j^P}$ and $\frac{\hat{\sigma}_j^I}{\hat{\varrho}_j^I}$ to zero in case they fall out of a pre-defined interval, which is in principle equivalent to applying simple chain ladder development factors assigned to the last few development years. This kind of truncation practice is followed in the present Bootstrap MCL calculations.

5.3.5 Semi-stochastic models

The following model is the modification of the one explained in section 4.4.3. Instead of addressing each triangle separately, consider the possibility of using other

	Boots. MCL Paid	MCL Paid	Realised Paid	Boots. MCL Incur.	MCL Incur.	Realised Incur.
1	0.0e+00	0.0e+00	0.0e+00	0.0e+00	0.0e+00	0.0e+00
2	-5.5e-06	2.1e+00	0.0e+00	-1.7e-04	6.2e-01	0.0e+00
3	7.9e-01	3.5e+00	0.0e+00	8.5e+01	1.5e+00	0.0e+00
4	2.1e+00	3.9e+26	0.0e+00	-7.7e+00	-1.2e+26	0.0e+00
5	6.1e+02	6.0e+25	1.2e+02	-6.0e+01	-1.8e+25	2.7e+01
6	8.4e+02	1.9e+26	1.6e+02	-8.7e+01	-5.7e+25	-2.0e+01
7	4.4e+02	5.6e+26	1.8e+02	2.5e+01	-1.7e+26	1.5e+01
8	6.4e+02	5.3e+26	3.5e+02	1.2e+01	-1.6e+26	-6.0e+01
9	1.2e+03	8.0e+26	8.6e+02	2.2e+02	-2.4e+26	-1.4e+02
10	1.7e+03	9.9e+26	1.9e+03	3.8e+02	-3.0e+26	-5.1e+01
Total	5.4e+03	3.5e+27	3.6e+03	5.7e+02	-1.0e+27	-2.3e+02

Table 5–8: Paid and incurred estimates divided into accident years (bootstrapped MCL, original MCL and actual). Data used from commercial auto, group code 8079.

companies' data from a corresponding product group. Being able to do so may either reflect the perspective of a regulatory organisation with collected data from insurance institutions, or data made publicly available voluntarily by the insurance institutions for collective improvement purposes. Eventually, the NAIC database is an example of the latter. The principle is similar to the one suggested in [36], however, instead of sampling from $\frac{C_{1,j+1}}{C_{1,j}}, \dots, \frac{C_{I-j,j+1}}{C_{I-j,j}}$ in one stand-alone run-off triangle, the new version is as follows.

Assumption 5.3 (Collective Semi-Stochastic) (1') Suppose that each subsequent cumulative claim has a multiplicative link to the previous one in accident year j through a random variable a_j. (2') a_j random variables are discrete uniform on the $\left\{ a_j(k) = \frac{\sum_{l=1}^{I-j} C_{l,j+1}^{(k)}}{\sum_{l=1}^{I-j} C_{l,j}^{(k)}} \;:\; k \in \{1, \dots, n\} \right\}$ set of development factors.

The assumption is similar to [36], however, now the cumulative claims are driven recursively by a_j random variables stemming from an unknown distribution, identically distributed across the run-off triangles.

Method 5.2 (Collective Semi-stochastic)

Step 1 Calculate chain ladder link ratios for $\mathcal{D}_1, \dots, \mathcal{D}_n$; $a_{j,k} \quad j \in \{1, \dots, I-1\}, k \in \{1, \dots, n\}$.

Step 2 For each j sample from $a_{j,1}, \dots, a_{j,n}$ with replacement; $a'_{j,1}, \dots, a'_{j,M}$, where M stands for an arbitrarily large sample size. $M = 5000$ in the actual examples in section 5.4.

Step 3 Multiplication of last cumulative observations: get the randomly generated ultimate claims. For a fixed company, $\hat{C}_{i,J}^{(s)} = C_{i,J-i} \prod_{j=J-i}^{J-1} a'_{j,s}$, $s \in \{1, \dots, M\}$.

abbreviation	*model*	*subsection*
boot.gamma	bootstrap model with gamma distribution	5.3.1
boot.od.pois	bootstrap model with overdispersed Poisson distr.	5.3.1
bootstrap.munich	Munich Chain Ladder with bootstrapping	5.3.4
CCL	correlated chain ladder model	5.3.2
CIT	correlated incremental trend model	5.3.2
cred.bootstrap.od.pois	credibility bootstrap with overdispersed Poisson	5.3.3
munich	Munich Chain Ladder (original)	5.3.4
SemiSt	collective semi-stochastic model	5.3.5

Table 5–9: Legends of reserving models.

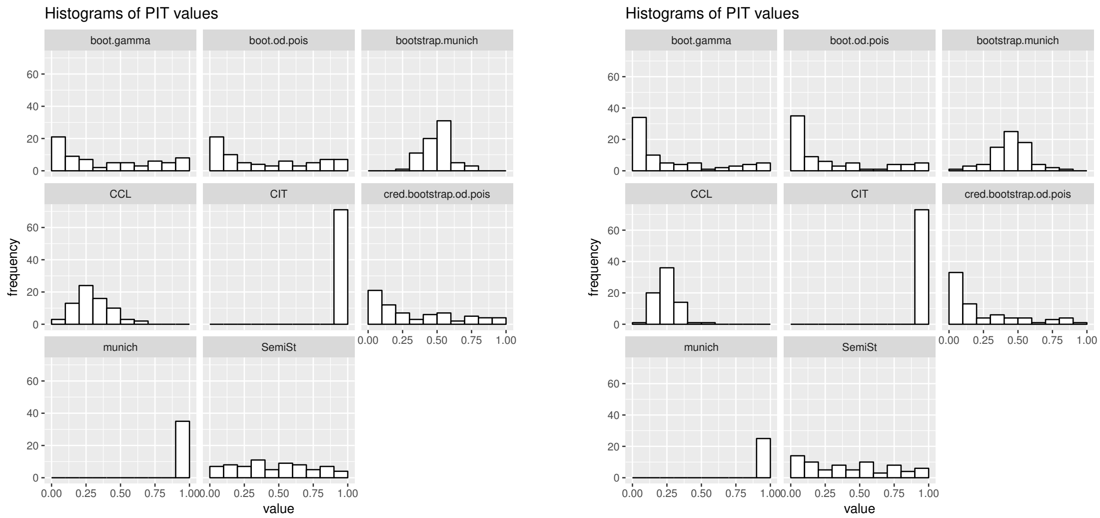
5.4 Comparing forecasts

Two out of the six sets of homogeneous risk groups available from NAIC are used to demonstrate results and draw conclusions. Commercial auto and private passenger auto liability data have been selected, due to the higher sample size (158 and 146 companies). Recall that the two samples still contain closely degenerate run-off triangles (almost all zero elements), which had to be filtered out in order to work with institutions for which all the reserving models provide meaningful results. Therefore, sample sizes have been reduced to 71 and 73. The only exception is the Munich Chain Ladder method, which is applicable to even less claim histories and would have rarefied the observations substantially. In each calculation, the actual sizes are indicated. Continuous ranked probability score, coverage and average width cannot be applied for the original MCL results. Metrics introduced in section 4.5 are quantified.

5.4.1 Probability integral transform

Each figure in the following subsections uses consistent abbreviations to indicate reserving methods, see Table 5–9.

Regardless of the question whether the stochastic method has a distribution or it is distribution-free, the empirical predictive distribution can always be generated by drawing randomly or bootstrapping a sufficient amount of samples. For a fixed reserving method, each quadrangle is associated with one $\hat{F}_i$ and the combination of these is used for backtesting. Results of the two business lines in Figure 5–2 suggest nearly identical inferences. It becomes instantly obvious that none of the reserving models provide unbiased estimation of the ultimate claim. In fact, the question is what exactly goes wrong with each one of them.



(a) **(b)**

Figure 5–2: (a) Histograms of PIT values from the commercial auto data. (b) Histograms of PIT values from the private passenger auto liability data.

The Munich chain ladder (MCL) is an odd one out, the only model discussed in the present chapter which is not suitable for producing a predictive distribution, and works only for a fraction of underlying run-off triangles, thus a lower amount of frequencies. Since MCL results in one single $\hat{UC}_{1,i}$ prediction, the $\hat{F}_i(z) = \begin{cases} 1, & z > \hat{UC}_{1,i} \\ 0, & \text{otherwise} \end{cases}$ frequencies are reflected on the MCL histograms. Besides, both related histograms prove that in each case, MCL consistently underestimated the actual outcome. The correlated incremental trend (CIT) model has a similar deficiency, resulting in under-dispersed predictions with one-sided biasedness.

The bootstrapped version of MCL and correlated chain ladder (CCL) models are both on the overdispersed spectrum. The former tends to result in a symmetric PIT histogram, suggesting that the expected value of the ultimate claim forecast is close to the expectation from the real distribution, which implies a significant improvement compared to the original MCL. PIT values of CCL model are biased to the left, as a sign of underestimation of ultimate claims.

The third group having similar results consists of bootstrap gamma and overdispersed Poisson and credibility bootstrap overdispersed Poisson models, having ∪-shaped PIT, i.e. narrow prediction intervals. Furthermore, biasedness can be observed to the left, indicating an underestimation of the real ultimate claims. The collective semi-stochastic approach performs relatively well in terms of PIT uniformity. We may conclude that the latter four models have the best qualities from a PIT perspective.

5.4.2 Continuous ranked probability score

On a set of observations and corresponding predictive distributions, the goal is to maximise the mean score, resulting in a ranking of competing predictive models through maximising the expected utility:

$$\mathcal{S}^{\text{model}} = \frac{1}{n}\sum_{i=1}^{n} S(P^{\text{model}}_{i\text{th company}}, x^{i\text{th company}}). \tag{5.11}$$

Let $P^{\text{model}}_{i\text{th company}} = \hat{F}_j = P^{\text{model}}_{i\text{th company}}\left(\hat{UC}_{1,i}, \ldots, \hat{UC}_{M,i}\right)$ stand for the empirical predictive distribution derived for company i on the basis of a fixed reserving model, where $\hat{UC}_{k,i}$ denotes the kth randomly generated ultimate claim for company i ($i = 1, \ldots, n$). Analytical formulas can rarely be derived for CRPS, not to mention the practical models of claims prediction, although, it is feasible if the distribution F is normal, see [46]. A reasonable question is how sensitively the mean score is exposed to extremely inappropriate models, i.e. if the sample size is relatively low and an outstanding score value is involved. For that reason the complete scale of score outcomes is proposed to be analysed in the form of a boxplot, the $-\log(-score)$ plotted for the sake of better visual understanding, see Figure 5–3. The higher the boxplot, the better the performance of forecast according to the scoring rule.

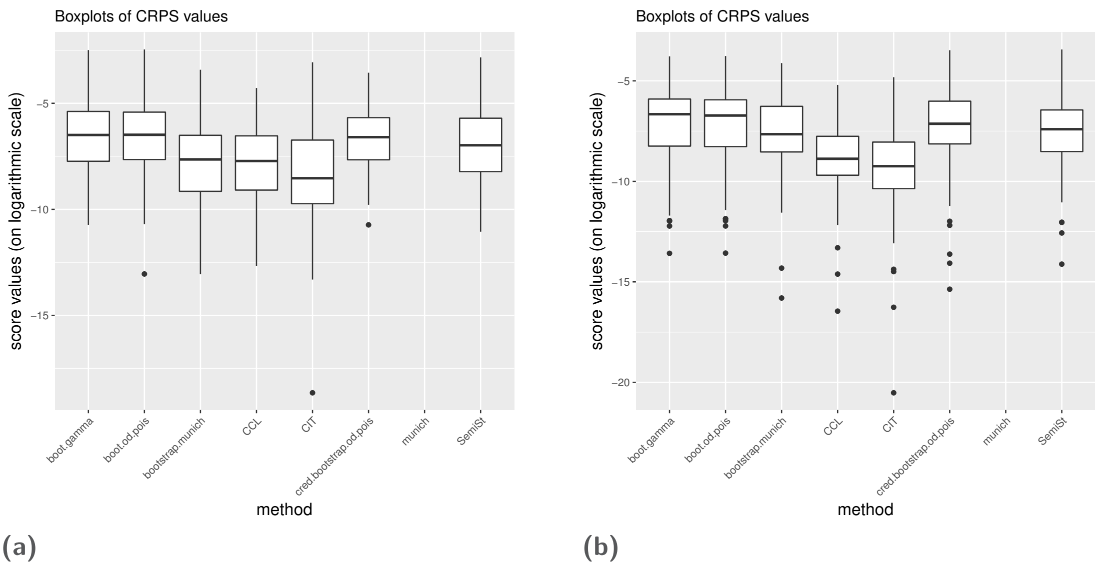


Figure 5–3: (a) Boxplots of CRPS values from the commercial auto data (after $-\log(-score)$–transformation). (b) Boxplots of CRPS values from the private passenger auto liability data (after $-\log(-score)$–transformation).

CRPS is not defined in relation to the MCL model due to the lack of predictive distribution. In Table 5–10 and 5–11 the mean CRPS values are demonstrated, which determine the ranking of competing models. In order to see whether an extreme value has influenced the mean outcome (defined in Equation 5.11), the median scores are added to the second column. Reserve calculations in accordance with the CIT model on both commercial and private passenger portfolios show scores of outstandingly large absolute value, implying that forecasts on some of the companies performed poorly.

In the calculation on the commercial auto data, the best performing model has been the credibility bootstrap overdispersed Poisson one, using experience ratemaking, whilst applied on the private passenger auto data it has performed worse than the other bootstrap methods. The semi-stochastic claims reserving technique becomes the third one applied on each of the data sets. Bootstrap MCL and CCL can be ranked behind these four models, and the CIT model yields significantly lower mean score values than the previous ones.

	Mean.CRPS	Median.CRPS	SampleSize
CIT	-1805000	-5082	71
CCL	-11880	-2260	71
boot.gamma	-2990	-662	71
boot.od.pois	-9404	-655	71
munich			0
bootstrap.munich	-20970	-2094	71
SemiSt	-4573	-1073	71
cred.bootstrap.od.pois	-2698	-733	71

Table 5–10: Average and median CRPS values from the commercial auto data.

	Mean.CRPS	Median.CRPS	SampleSize
CIT	-11410000	-10350	73
CCL	-247900	-7163	73
boot.gamma	-22620	-780	73
boot.od.pois	-23200	-831	73
munich			0
bootstrap.munich	-132800	-2108	73
SemiSt	-31760	-1644	73
cred.bootstrap.od.pois	-101400	-1253	73

Table 5–11: Average and median CRPS values from the private passenger auto liability data.

5.4.3 Coverage and average width

In the calculations with NAIC data, each width in the average calculation formula defined in section 4.5.3 is normalised in every triangle with the realised IBNR value. The normalising value stands for the lower triangle sum in case of an incremental point of view, or, in other words, the ultimate claim reduced by the payment already available in the upper triangle. Hence, it reflects the average span interval as a unit of realised IBNR value.

In the ideal case of coinciding predictive and actual probability measures $P \stackrel{d}{=} Q$, coverage α equals to α for any given $\alpha \in (0, 1)$. Tables 5–12 and 5–13 calculated on the basis of two α values prove that the applied models produce coverages that are far from ideal. The original MCL method does not have any coverage or average width output due to lack of predictive distribution. CIT and bootstrap MCL show the most inappropriate characteristics, in essence with degenerate coverages, either equal or close to 0 or 1. CCL performs better in the sense that the lower $\alpha = 67\%$ coverage is 84% and 94% in the two cases. The credibility bootstrap and original bootstrap gamma and overdispersed Poisson methods result in similar coverage and average width: Measures are balanced among these three models, and have the narrowest sharpness. The collective semi-stochastic method results in coverages closest to identity, however, at the cost of having wider average width values.

	67% cover	90% cover	67% width	90% width	SampleSize
CIT	0.00	0.00	0.01	0.02	71
CCL	0.84	1.00	4.96	10.60	71
boot.gamma	0.45	0.79	1.04	2.43	71
boot.od.pois	0.45	0.78	1.00	2.13	71
munich	0.00	0.00	0.00	0.00	56
bootstrap.munich	1.00	1.00	37.83	113.90	71
SemiSt	0.73	0.99	1.51	3.55	71
cred.bootstrap.od.pois	0.51	0.75	1.13	2.34	71

Table 5–12: Coverage and average width from the commercial auto data.

	67% cover	90% cover	67% width	90% width	SampleSize
CIT	0.00	0.00	0.00	0.01	73
CCL	0.94	1.00	5.38	10.30	73
boot.gamma	0.30	0.59	0.59	1.14	73
boot.od.pois	0.32	0.57	0.58	1.12	73
munich	0.00	0.00	0.00	0.00	61
bootstrap.munich	0.97	1.00	98.43	411.20	73
SemiSt	0.59	0.93	0.97	2.33	73
cred.bootstrap.od.pois	0.37	0.59	0.58	1.03	73

Table 5–13: Coverage and average width from the private passenger auto liability data.

5.4.4 Mean square error of prediction

Results calculated here differ from the original definition in the sense that each outcome is normalised by the ultimate reserve. The reason corresponds to the one discussed in Section 5.2, i.e. the magnitudinal discrepancies among the claims in distinct companies. Hence, instead of $E\left((\xi_i - \eta_i)^2|\mathcal{D}_i\right)$ we estimate $E\left((\frac{\eta_i}{\xi_i} - 1)^2|\mathcal{D}_i\right)$. Draw a random sample from the distribution of η_i determined by the forecasting model, and the real observed realisation of ξ_i; $\hat{UC}_{1,i}, \ldots, \hat{UC}_{M,i}$ and UC_i.

Proposition 5.2 $\frac{1}{M}\sum\limits_{j=1}^{M}\frac{(\hat{UC}_{j,i}-UC_i)^2}{UC_i^2}$ *is an unbiased estimator of* $E\left((\frac{\eta_i}{\xi_i} - 1)^2|\mathcal{D}_i\right)$.

Proof.

$$E\left(\frac{1}{M}\sum_{j=1}^{M}\frac{(\hat{UC}_{j,i} - UC_i)^2}{UC_i^2}|\mathcal{D}_i\right) = E\left(\frac{(\hat{UC}_{1,i} - UC_i)^2}{UC_i^2}|\mathcal{D}_i\right) = E\left(\frac{(\eta_i - \xi_i)^2}{\xi_i^2}|\mathcal{D}_i\right). \tag{5.12}$$

■

Finally, the MSEP estimator of the model, unconstrained on the upper triangle is the average of the elements calculated for each company i. However, should the mean be dominated by any extreme value, the median of conditional MSEPs is included in the calculation results. Observe the differing values in Table 5–14 and 5–15, supporting the actuary with insufficient background in order to determine reliable methods on the datasets. Taking exclusively the MSEP into account in model decisions is clearly not the proper way of ranking them and does not provide information concerning the appropriateness of predictive distribution.

	Mean.Msep	Median.Msep	SampleSize
CIT	127.7	1.0	71
CCL	445.1	6.2	71
boot.gamma	352.6	0.2	71
boot.od.pois	6137.0	0.1	71
munich	1.9	0.0	52
bootstrap.munich	6235000.0	16.5	71
SemiSt	4.3	1.7	71
cred.bootstrap.od.pois	3112.0	0.2	71

Table 5–14: Mean square error of prediction from the commercial auto data.

5.4.5 Ranking Algorithm

We summarise the algorithmic steps of the ranking framework. Suppose that the triangles stem from one homogeneous risk group.

	Mean.Msep	Median.Msep	SampleSize
CIT	61800000.0	1.0	73
CCL	25.9	12.9	73
boot.gamma	38450.0	0.1	73
boot.od.pois	874.0	0.1	73
munich	2.1	0.0	59
bootstrap.munich	2791000.0	3.1	73
SemiSt	14.0	6.5	73
cred.bootstrap.od.pois	7.7	0.1	73

Table 5–15: Mean square error of prediction from the private passenger auto liability data.

I | *Stochastic forecast phase.* For $meth \in \{$ bootstrap gamma, bootstrap ODP, $\dots\}$, for $j \in \{$set of companies$\}$, generate M ultimate claim values.
Result: $\hat{UC}_{1,j,meth}, \dots, \hat{UC}_{M,j,meth}$ $\forall j$ $\forall meth$.

II | *Backtest phase.* For $meth \in$ $\{$ bootstrap gamma, bootstrap ODP, $\dots\}$, $j \in$ $\{$set of companies$\}$ calculate PIT, CRPS, coverage, sharpness, MSEP from $\hat{UC}_{1,j,meth}$, $\dots$, $\hat{UC}_{M,j,meth}$ and real UC_j.
Result: (a) $PIT_{j,meth} \in (0,1)$, (b) $CRPS_{j,meth} \in \mathbb{R}_-$, (c) $cover_{j,meth,p} \in (0,1)$, (d) $sharp_{j,meth,p} \in \mathbb{R}_+$, (e) $MSEP_{j,meth} \in \mathbb{R}_+$ $\forall j$ $\forall meth$ $\forall p \in \{67\%, 90\%\}$.

III | *Ranking phase.* Separate comparison of metrics (a)-(e). Combined comparison of metrics (f). (We assume to compare 7 stochastic methods, excluding MCL.)

a) Calculate the entropy $PIT_{\cdot,meth_i}$ of each set $\{PIT_{j,meth} : \forall j\}$ and order $PIT_{\cdot,meth_1} > \dots > PIT_{\cdot,meth_7}$. Assign rank i to $meth_i$, the lower the rank the better the performance.
b) Calculate average CRPS and order $CRPS_{\cdot,meth_1} > \dots > CRPS_{\cdot,meth_7}$. Assign rank i to $meth_i$.
c) Calculate coverage values $cover_{\cdot,meth_i,p}$ and order $(cover_{\cdot,meth_1,p} - p)^2 < \dots < (cover_{\cdot,meth_7,p} - p)^2$ for each p and assign rank i to $meth_i$. For each method, take the arithmetic average of the two ranks.
d) Calculate sharpness values $sharp_{\cdot,meth_i,p}$ and order $sharp_{\cdot,meth_1,p} < \dots < sharp_{\cdot,meth_7,p}$ for each p and assign rank i to $meth_i$. Similarly to coverage take the average of the two ranks for each method.
e) Calculate MSEP values and rank as for sharpness.
f) For $meth \in \{$ bootstrap gamma, bootstrap ODP, $\dots\}$ determine $rank^{total}_{meth_i} = rank^{PIT}_{meth_i} + rank^{CRPS}_{meth_i} + rank^{cover}_{meth_i} + rank^{sharp}_{meth_i} + rank^{MSEP}_{meth_i}$. Method k performs better than l if $rank^{total}_{meth_k} < rank^{total}_{meth_l}$.

Observe that the metrics have identical weights in ranking, which is an arbitrary choice. These steps describe a combined ranking based on different characteristics.

However, this ranking should not be applied without scrutinising PIT, CRPS, etc. separately in order to see the exact weakness of a reserving method. The ranking results per business line can be found in Table 5–16. Observe that in contrast to all other models, the bootstrap gamma one never ranked worse than 3.

	comauto	medmal	ppauto	prodliab	wkcomp	othliab
CIT	5	4	7	5	7	6
CCL	6	6	5	6	5	4
boot.gamma	2	1	3	1	2	1
boot.od.pois	4	3	4	2	3	2
bootstrap.munich	7	7	6	7	6	7
SemiSt	1	5	2	3	1	3
cred.bootstrap.od.pois	3	2	1	4	4	5

Table 5–16: Combined rankings of reserving methods per business line. (Excl. MCL.)

5.5 Conclusion

Rapidly increasing computational power has led to a shift from deterministic claims reserving models to stochastic ones. Simultaneously, the validation of model appropriateness has received proportionally less attention from researchers. In our view it is crucial to understand the suitability of different methods for the calculation of remaining future payments in an insurance portfolio, and to compare them from several perspectives. We have interpreted claims reserving as a probabilistic forecast. Data sets of six business lines from American insurance institutions supported calculations in order to remain in contact with real-life claim outcomes. We would have also used European ones if there were any at our disposal.

Eight different models have been introduced with key parameter estimation details, out of which five principally different method families can be distinguished, one including MCMC simulations. Two of the models were first introduced in [72], using not only the individual insurers', but collective claims observations from other companies for calibration, see experience ratemaking embedded into the credibility bootstrap overdispersed Poisson model. Semi-stochastic and credibility bootstrap models have been among the best performing ones, however, results lack significant evidence that these would considerably outperform their regular bootstrap counterparts.

The primary objective of the present chapter was to improve the decision making among several available models applied on run-off triangles, by defining and calculating measures of the actual and predictive distributions. Given that actual distributions can hardly be extracted, we have used empirical distributions from real ultimate claims data. Goodness-of-fit measures of predictive distribution are clearly more informative than exclusively observing the mean square error of the prediction. Probability integral transform is better than Kolmogorov–Smirnov or Cramér–von Mises in the sense that it highlights what goes wrong with the hypothesis. CRPS can widely be applied on distributions with no constraint on absolute continuity, defining a ranking among competing models. Further characteristics such as coverage and sharpness explain the central prediction interval and its expected width. Methods with bootstrapping have shown the best performance in general, along with the semi-stochastic model.

Appendices

A Scores excluded from section 2.5

Besides the definitions discussed in section 2.5, several further concepts can be found in literature, such as the Beta family, Winkler's score, etc. The application of the chosen scores is justified by the discrete probabilistic behaviour of claim numbers.

Definition 0.4 (spherical score) Let $\alpha > 1$ be real. The pseudospherical score is defined as $S(p,i) = \frac{p_i^{\alpha-1}}{\left(\sum_j p_j^\alpha\right)^{\frac{\alpha-1}{2}}}$. If $\alpha = 2$, the score is the spherical score.

The entropy function is $G(p) = \left(\sum_{i\in|\Omega|} p_i^\alpha\right)^{1/\alpha}$ and the associated Bregman divergence is $d(p,q) = \left(\sum_{i\in|\Omega|} q_i^\alpha\right)^{1/\alpha} - \frac{\sum_{i\in|\Omega|} p_i q_i^{\alpha-1}}{\left(\sum_{i\in|\Omega|} q_i^\alpha\right)^{(\alpha-1)/\alpha}}$

Definition 0.5 (zero-one score) Let $M(p) = \{i : p_i = \max p_j\}$ denote the modes of p. The zero-one score is defined by $S(p,i) = \begin{cases} \frac{1}{\#M(p)}, i \in M(p) \\ 0, \text{otherwise} \end{cases}$.

The entropy function is $G(p) = \max_{i\in|\Omega|} p_i$ and the divergence function is $d(p,q) = \max_{i\in|\Omega|} q_i - \frac{\sum_{i\in M(p)} q_i}{\#M(p)}$.

For the sake of completeness, one has to apply other measures in case of random variables of uncountable range. [40, 70] address claim severities. For more details of density or distribution function forecasts in general see [46, 47]. For density forecasts, consider L_α probability measures on $(\Omega, \mathcal{A})$ and let $p \ll \mu$, i.e. the μ-density p exists. Let the norm $\|p\|_\alpha = (\int p(\omega)^\alpha \mu(d\omega))^{1/\alpha}$ be defined. The peers of the scores already introduced, corresponding to absolutely continuous distributions, can be defined as follows.

- Quadratic score: $S(p,\omega) = 2p(\omega) - \|p\|_2^2$.

- Logarithmic score: $S(p,\omega) = \log p(\omega)$. (The limiting case of the pseudospherical score as $\alpha \to 1$.)
- Spherical score: $S(p,\omega) = p(\omega)/\left\|p\right\|_2$.

The spectrum of proper scoring rules applied is significantly wider than the ones enumerated in this section, among others, the so-called kernel scores, scoring rules for interval forecasts, the BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) scores, which are likelihood-based and relate to the logarithmic score.

B Transition matrices of bonus–malus systems

B.13 shows the transition probability matrix of the Hungarian bonus–malus system as a function of claims frequency. Element (i,j) stands for the transition probability from premium level C_{16-i} to C_{16-j}. Recall that C_1 is the class with the highest bonus and C_{15} is the one with the highest penalty.

$$M(\lambda) = \begin{pmatrix}
1-e^{-\lambda} & e^{-\lambda} & 0 & \dots & 0 & 0 & 0 \\
1-e^{-\lambda} & 0 & e^{-\lambda} & \ddots & 0 & 0 & 0 \\
1-e^{-\lambda} & 0 & 0 & \ddots & 0 & 0 & 0 \\
1-e^{-\lambda} & \lambda e^{-\lambda} & 0 & \ddots & 0 & 0 & 0 \\
1-(\lambda+1)e^{-\lambda} & 0 & \lambda e^{-\lambda} & \ddots & 0 & 0 & 0 \\
1-(\lambda+1)e^{-\lambda} & \lambda^2 \cdot e^{-\lambda}/2! & 0 & \ddots & 0 & 0 & 0 \\
1-(\lambda^2/2!+\lambda+1)e^{-\lambda} & 0 & \lambda^2 \cdot e^{-\lambda}/2! & \ddots & 0 & 0 & 0 \\
1-(\lambda^2/2!+\lambda+1)e^{-\lambda} & \lambda^3 \cdot e^{-\lambda}/3! & 0 & \ddots & 0 & 0 & 0 \\
1-(\lambda^3/3!+\lambda^2/2!+\lambda+1)e^{-\lambda} & 0 & \lambda^3 \cdot e^{-\lambda}/3! & \ddots & 0 & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
1-(\lambda^3/3!+\lambda^2/2!+\lambda+1)e^{-\lambda} & 0 & 0 & \ddots & 0 & 0 & e^{-\lambda} \\
1-(\lambda^3/3!+\lambda^2/2!+\lambda+1)e^{-\lambda} & 0 & 0 & \dots & \lambda e^{-\lambda} & 0 & e^{-\lambda}
\end{pmatrix} \tag{B.13}$$

C Run-off triangles

	1	2	3	4	5	6
1	126 747 199	601 228 543	297 403 379	165 198 453	680 414 968	405 097 366
2	62 428 642	451 138 014	460 693 799	155 646 520	322 900 614	160 950 461
3	53 710 018	582 525 687	376 291 743	262 637 270	254 425 516	261 681 230
4	92 060 095	587 566 179	528 869 566	313 577 319	446 860 243	116 489 238
5	64 411 976	765 398 108	483 444 263	479 496 593	1 704 359 785	180 683 591
6	90 307 344	645 283 229	748 848 104	248 603 703	292 906 556	152 940 157

Table C–17: Incremental loss triangle public payments data completed with the lower part. (Accident year / Development year, without the exact accident years)

	1	2	3	4	5	6	7	8	9	10
1981	5 012	8 269	10 907	11 805	13 539	16 181	18 009	18 608	18 662	18 834
1982	106	4 285	5 396	10 666	13 782	15 599	15 496	16 169	16 704	
1983	3 410	8 992	13 873	16 141	18 735	22 214	22 863	23 466		
1984	5 655	11 555	15 766	21 266	23 425	26 083	27 067			
1985	1 092	9 565	15 836	22 169	25 955	26 180				
1986	1 513	6 445	11 702	12 935	15 852					
1987	557	4 020	10 946	12 314						
1988	1 351	6 947	13 112							
1989	3 133	5 395								
1990	2 063									

Table C–18: Cumulative loss triangle RAA. (Accident year / Development year)

	1	2	3	4	5	6	7	8	9	10	11
1977	153 638	342 050	476 584	564 040	624 388	666 792	698 030	719 282	735 904	750 344	762 544
1978	178 536	404 948	563 842	668 528	739 976	787 966	823 542	848 360	871 022	889 022	
1979	210 172	469 340	657 728	780 802	864 182	920 268	958 764	992 532	1 019 932		
1980	211 448	464 930	648 300	779 340	858 334	918 566	964 134	1 002 134			
1981	219 810	486 114	680 764	800 862	888 444	951 194	1 002 194				
1982	205 654	458 400	635 906	765 428	862 214	944 614					
1983	197 716	453 124	647 772	790 100	895 700						
1984	239 784	569 026	833 828	1 024 228							
1985	326 304	798 048	1 173 448								
1986	420 778	1 011 178									
1987	496 200										

Table C–19: Cumulative loss triangle ABC. (Accident year / Development year)

	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315		
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268			
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311				
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712					
7	440 832	1 288 463	2 419 861	3 483 130						
8	359 480	1 421 128	2 864 498							
9	376 686	1 363 294								
10	344 014									

Table C–20: Cumulative loss triangle GenIns. (Accident year / Development year)

Bibliography

[1]K. Antonio and R. Plat. "Micro-level stochastic loss reserving for general insurance". In: *Scandinavian Actuarial Journal* 2014.7 (2014), pp. 649–669. DOI: 10.1080/03461238.2012.755938. eprint: https://doi.org/10.1080/03461238.2012.755938 (cit. on p. 102).

[2]M. Arató and L. Martinek. "Comparison of Reserving and Spatial Models in Insurance". In: *Current Topics on Risk Analysis: ICRA6 and RISK 2015 Conference*. Ed. by M. Guillén, Á A Juan, H. Ramalhinho, I. Serra, and C. Serrat. Fundación MAPFRE, 2015, pp. 71–78 (cit. on pp. 4, 68, 85).

[3]M. Arató and L. Martinek. "Estimation of claim numbers in automobile insurance". In: *Annales Univ. Sci. Budapest., Sect. Comp.* 42.1 (2014), pp. 19–35 (cit. on pp. 4, 9, 29).

[4]M. Arató, L. Martinek, and M. Mályusz. "Simulation Based Comparison of Stochastic Claims Reserving Models in General Insurance". In: *Studia Scientiarum Mathematicarum Hungarica* 54.2 (2017), pp. 241–275. DOI: 10.1556/012.2017.54.2.1364 (cit. on pp. 4, 67, 68, 75, 79, 82, 84, 85, 104).

[5]F. Ashe. "An Essay at Measuring the Variance of Estimates of Outstanding Claim Payments". In: *Astin Bulletin* 16.S1 (1986), S99–S113. DOI: 10.1017/S0515036100011685 (cit. on p. 78).

[6]S. Asmussen. "Modeling and Performance of Bonus-Malus Systems: Stationarity versus Age-Correction". In: *Risks* 2.1 (2014), pp. 49–73. DOI: 10.3390/risks2010049 (cit. on p. 40).

[7]European Central Bank. *Statistics on euro area insurance corporations*. [Online; accessed 6-November-2018]. 2018. eprint: https://www.ecb.europa.eu/press/pdf/icb/ecb.icb2017q4.en.pdf?7f6cefed9e895b32bab91cf9772ca4d3 (cit. on p. 65).

[8]S. Baran, D. Nemoda, and A. Horányi. "Statistical post-processing of probabilistic wind speed forecasting in Hungary". In: *Meteorologische Zeitschrift* 22.3 (2013), pp. 273–282. DOI: 10.1127/0941-2948/2013/0428 (cit. on p. 85).

[9]G. L. Barnett and B. Zehnwirth. "Best estimates for reserves". In: *Proceedings of the Casualty Actuarial Society LXXXVII*. 2000 (cit. on p. 75).

[10]J. M. Bernardo. "Expected Information as Expected Utility". In: *The Annals of Statistics* 7.3 (1979), pp. 686–690. DOI: 10.1214/aos/1176344689 (cit. on p. 23).

[11]S. Björkwall, O. Hössjer, and E. Ohlsson. “Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving”. In: *Scandinavian Actuarial Journal* 2009.4 (2009), pp. 306–331. DOI: 10.1080/03461230903239738. eprint: https://doi.org/10.1080/03461230903239738 (cit. on pp. 67, 78).

[12]J.-P. Boucher, M. Denuit, and M. Guillén. “Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions”. In: *Variance* 2.1 (2008), pp. 135–162 (cit. on p. 11).

[13]G. W. Brier. “Verification of Forecasts Expressed in Terms of Probability”. In: *Monthly Weather Review* 78.1 (1950), pp. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2. eprint: https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2 (cit. on p. 24).

[14]M. Buchwalder, H. Bühlmann, M. Merz, and M. V. Wüthrich. “The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited)”. In: *Astin Bulletin* 36.02 (2006), pp. 521–542. DOI: 10.1017/S0515036100014628 (cit. on p. 88).

[15]H. Bühlmann. “Experience rating and credibility”. In: *Astin Bulletin* 4.03 (1967), pp. 199–207. DOI: 10.1017/S0515036100008989 (cit. on p. 109).

[16]H. Bühlmann. “Experience rating and credibility”. In: *ASTIN Bulletin* 5.2 (1969), pp. 157–165. DOI: 10.1017/S0515036100008023 (cit. on p. 109).

[17]H. Bühlmann and A. Gisler. *A course in credibility theory and its applications*. Springer Science & Business Media, 2006. DOI: 10.1007/3-540-29273-X (cit. on p. 111).

[18]A. Charpentier, A. David, and R. Elie. “Optimal Claiming Strategies in Bonus Malus Systems and Implied Markov Chains”. In: *Risks* 5.4 (2017), p. 58. DOI: 10.3390/risks5040058 (cit. on p. 62).

[19]P. F. Christoffersen. “Evaluating Interval Forecasts”. In: *International Economic Review* 39.4 (1998), pp. 841–862. DOI: 10.2307/2527341 (cit. on p. 85).

[20]G. Coene and L. G. Doray. “A Financially Balanced Bonus-Malus System”. In: *ASTIN Bulletin* 26.1 (1996), pp. 107–116. DOI: 10.2143/AST.26.1.563236 (cit. on p. 46).

[21]S. Cross and J. Doucette. “Measurement of asbestos bodily injury liabilities”. In: *Proceedings of the Casualty Actuarial Society*. Vol. 84. 1997, pp. 187–300 (cit. on p. 69).

[22]C. Czado, T. Gneiting, and L. Held. “Predictive Model Assessment for Count Data”. In: *Biometrics* 65.4 (2009), pp. 1254–1261. DOI: 10.1111/j.1541-0420.2009.01191.x (cit. on p. 83).

[23]A. P. Dawid. “Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach”. English. In: *Journal of the Royal Statistical Society. Series A (General)* 147.2 (1984), pp. 278–292. DOI: 10.2307/2981683 (cit. on pp. 81, 82).

[24] *Decree No. 21/2011 (VI. 10) of the Ministry for National Economy, Hungary.* 2011. eprint: https://net.jogtar.hu/jogszabaly?docid=a1100021.ngm (cit. on pp. 11, 53).

[25] P. Delaporte. "Quelques problèmes de statistique mathématique posés par l'assurance automobile et le bonus pour non sinistre". In: *Bulletin Trimestriel de l'Institut des Actuaires Français* 227 (1959), pp. 87–102 (cit. on p. 47).

[26] M. Denuit and J. Dhaene. "Bonus-Malus scales using exponential loss functions". In: *Blätter der DGVFM* 25.1 (2001), pp. 13–27. DOI: 10.1007/BF02857113 (cit. on p. 31).

[27] M. Denuit and P. Lambert. "Smoothed NPML estimation of the risk distribution underlying Bonus-Malus systems". In: *Proceedings of the Casualty Actuarial Society.* Vol. 88. 198. 2001, p. 142 (cit. on p. 11).

[28] M. Denuit, X. Maréchal, S. Pitrebois, and J-F. Walhin. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems.* John Wiley & Sons, 2007. DOI: 10.1002/9780470517420 (cit. on pp. 7, 10, 46).

[29] F. X. Diebold, T. A. Gunther, and A. S. Tay. "Evaluating density forecasts with Applications to Financial Risk Management". In: *International Economic Review.* Technical Working Paper Series 39 (1998), pp. 863–883. DOI: 10.3386/t0215 (cit. on pp. 81, 82).

[30] G. Dionne and C. Vanasse. "A generalization of automobile insurance rating models: the negative binomial distribution with a regression component". In: *Astin Bulletin* 19.2 (1989), pp. 199–212. DOI: 10.2143/AST.19.2.2014909 (cit. on p. 8).

[31] B. Efron. "Bootstrap Methods: Another Look at the Jackknife". In: *The Annals of Statistics* 7.1 (Jan. 1979), pp. 1–26. DOI: 10.1214/aos/1176344552 (cit. on p. 77).

[32] B. Efron. "Computers and the Theory of Statistics: Thinking the Unthinkable". In: *SIAM Review* 21.4 (1979), pp. 460–480. DOI: 10.1137/1021092 (cit. on p. 77).

[33] P. D. England and R. J. Verrall. "Analytic and bootstrap estimates of prediction errors in claims reserving". In: *Insurance: Mathematics and Economics* 25.3 (1999), pp. 281–293. DOI: 10.1016/S0167-6687(99)00016-5 (cit. on p. 78).

[34] P. D. England and R. J. Verrall. "Predictive Distributions of Outstanding Liabilities in General Insurance". In: *Annals of Actuarial Science* 1.02 (Aug. 2006), pp. 221–270. DOI: 10.1017/S1748499500000142 (cit. on p. 74).

[35] P.D. England and R.J. Verrall. "Stochastic Claims Reserving in General Insurance". In: *British Actuarial Journal* 8.3 (2002), pp. 443–518. DOI: 10.1017/S1357321700003809 (cit. on pp. 2, 65, 66, 72–75, 78, 92).

[36] T. Faluközy, I. I. Vitéz, and M. Arató. "Stochastic models for claims reserving in insurance business". In: *Recent Advances In Stochastic Modeling And Data Analysis.* World Scientific, 2007. Chap. 13, pp. 102–113. DOI: 10.1142/9789812709691_0013 (cit. on pp. 80, 118).

[37]FBTO. *Schadevrije jaren en no-claimkorting autoverzekering*. https://www.fbto.nl/documenten/Tabel_Schadevrije_jaren_internet.pdf. [Online; accessed 9-March-2018]. 2018 (cit. on pp. 11, 53).

[38]C. A. T. Ferro. “Comparing Probabilistic Forecasting Systems with the Brier Score”. In: *Weather and Forecasting* 22.5 (2007), pp. 1076–1088. DOI: 10.1175/WAF1034.1 (cit. on p. 24).

[39]G. Fishman. *Monte Carlo. Concepts, Algorithms, and Applications*. Springer Series in Operations Research and Financial Engineering. Springer Science & Business Media, 1996. DOI: 10.1007/978-1-4757-2553-7 (cit. on p. 10).

[40]N. E. Frangos and S. D. Vrontos. “Design of Optimal Bonus-Malus Systems With a Frequency and a Severity Component On an Individual Basis in Automobile Insurance”. In: *ASTIN Bulletin* 31.1 (2001), pp. 1–22. DOI: 10.2143/AST.31.1.991 (cit. on pp. 9, 129).

[41]E. W. Frees. *Regression modeling with actuarial and financial applications*. International Series on Actuarial Science. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511814372 (cit. on p. 7).

[42]S. Frühwirth-Schnatter and S. Pyne. “Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions”. In: *Biostatistics* 11.2 (2010), pp. 317–336. DOI: 10.1093/biostatistics/kxp062. eprint: /oup/backfile/Content_public/Journal/biostatistics/11/2/10.1093/biostatistics/kxp062/2/kxp062.pdf (cit. on p. 108).

[43]M. Gesmann. *Statistical Methods and Models for Claims Reserving in General Insurance*. https://cran.r-project.org/web/packages/ChainLadder/ChainLadder.pdf. [Online; accessed 10-October-2018]. May 2018 (cit. on p. 92).

[44]A. Gisler and M. V. Wüthrich. “Credibility for the chain ladder reserving method”. In: *Astin Bulletin* 38.02 (2008), pp. 565–600. DOI: 10.1017/S0515036100015294 (cit. on pp. 110, 111).

[45]T. Gneiting. “Making and Evaluating Point Forecasts”. In: *Journal of the American Statistical Association* 106.494 (2011), pp. 746–762. DOI: 10.1198/jasa.2011.r10138 (cit. on pp. 1, 21).

[46]T. Gneiting and A. E. Raftery. “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477 (2007), pp. 359–378. DOI: 10.1198/016214506000001437. eprint: https://doi.org/10.1198/016214506000001437 (cit. on pp. 1, 21, 23, 67, 121, 129).

[47]T. Gneiting, F. Balabdaoui, and A. E. Raftery. “Probabilistic forecasts, calibration and sharpness”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.2 (2007), pp. 243–268. DOI: 10.1111/j.1467-9868.2007.00587.x (cit. on pp. 3, 67, 81, 82, 86, 88, 129).

[48]E. Gómez-Déniz. "Bivariate credibility bonus–malus premiums distinguishing between two types of claims". In: *Insurance: Mathematics and Economics* 70 (2016), pp. 117–124. DOI: 10.1016/j.insmatheco.2016.06.009 (cit. on pp. 31, 54).

[49]I. J. Good. "Rational decisions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1952), pp. 107–114 (cit. on p. 24).

[50]S. Gschlössl and C. Czado. "Spatial modelling of claim frequency and claim size in non-life insurance". In: *Scandinavian Actuarial Journal* 2007.3 (2007), pp. 202–225. DOI: 10.1080/03461230701414764. eprint: https://doi.org/10.1080/03461230701414764 (cit. on p. 84).

[51]T. M. Hamill. "Interpretation of Rank Histograms for Verifying Ensemble Forecasts". In: *Monthly Weather Review* 129.3 (2001), pp. 550–560. DOI: 10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2. eprint: https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2 (cit. on p. 82).

[52]H. Hersbach. "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems". In: *Weather and Forecasting* 15.5 (2000), pp. 559–570. DOI: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2. eprint: http://dx.doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2 (cit. on p. 85).

[53]C.-A. S. Staël von Holstein. "A Family of Strictly Proper Scoring Rules Which Are Sensitive to Distance". In: *Journal of Applied Meteorology* 9.3 (1970), pp. 360–364. DOI: 10.1175/1520-0450(1970)009<0360:AFOSPS>2.0.CO;2 (cit. on p. 23).

[54]P. de Jong and G. Z. Heller. *Data Sets*. http://www.businessandeconomics.mq.edu.au/our_departments/Applied_Finance_and_Actuarial_Studies/research/books/GLMsforInsuranceData/data_sets. [Online; accessed 18-March-2018]. 2008 (cit. on p. 31).

[55]V. Jürgens, S. Ess, T. Cerny, and P. Vounatsou. "A Bayesian generalized age–period–cohort power model for cancer projections". In: *Statistics in Medicine* 33.26 (2014), pp. 4627–4636. DOI: 10.1002/sim.6248 (cit. on p. 22).

[56]G. Kapetanios, J. Mitchell, S. Price, and N. Fawcett. "Generalised density forecast combinations". In: *Journal of Econometrics* 188.1 (2015), pp. 150–165. DOI: 10.1016/j.jeconom.2015.02.047 (cit. on p. 22).

[57]D. Karlis, G. Tzougas, and N. Frangos. "Confidence intervals of the premiums of optimal bonus malus systems". In: *Scandinavian Actuarial Journal* 2018.2 (2018), pp. 129–144. DOI: 10.1080/03461238.2017.1307267 (cit. on p. 47).

[58]N. Klein, M. Denuit, S. Lang, and T. Kneib. "Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape". In: *Insurance: Mathematics and Economics* 55 (2014), pp. 225–249. DOI: 10.1016/j.insmatheco.2014.02.001 (cit. on p. 84).

[59]J. Lemaire. *Automobile Insurance. Actuarial Models*. second. Kluwer–Nijhoff, 1985. DOI: 10.1007/978-94-015-7708-3 (cit. on pp. 1, 7, 8).

[60] J. Lemaire. *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, 1995. DOI: 10.1007/978-94-011-0631-3 (cit. on pp. 8, 52).

[61] J. Lemaire. "Bonus-Malus Systems: The European and Asian Approach to Merit-Rating". In: *North American Actuarial Journal* 2.1 (1998), pp. 26–38. DOI: 10.1080/10920277.1998.10595668 (cit. on pp. 1, 7, 12, 60, 62).

[62] J. Lemaire and H. Zi. "A Comparative Analysis of 30 Bonus-Malus Systems". In: *ASTIN Bulletin: The Journal of the IAA* 24.2 (1994), pp. 287–309. DOI: 10.2143/AST.24.2.2005071 (cit. on pp. 11, 45, 53, 60).

[63] J. Lemaire, S. C. Park, and K. C. Wang. "The impact of covariates on a bonus–malus system: an application of Taylor's model". In: *European Actuarial Journal* 5.1 (2015), pp. 1–10. DOI: 10.1007/s13385-015-0107-6 (cit. on p. 31).

[64] W. K. Leong, S. S. Wang, and H. Chen. "Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data". In: *Variance* 8.2 (2014), pp. 182–202 (cit. on pp. 78, 103).

[65] S. Lerch, T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting. "Forecaster's Dilemma: Extreme Events and Forecast Evaluation". In: *Statist. Sci.* 32.1 (Feb. 2017), pp. 106–127. DOI: 10.1214/16-STS588 (cit. on p. 22).

[66] H. Liu and R. J. Verrall. "Bootstrap estimation of the predictive distributions of reserves using paid and incurred claims". In: *Variance* 4 (2010), pp. 121–135 (cit. on p. 115).

[67] K. Loimaranta. "Some asymptotic properties of bonus systems". In: *ASTIN Bulletin* 6.3 (1972), pp. 233–245. DOI: 10.1017/S0515036100011028 (cit. on p. 38).

[68] T. Mack. "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates". In: *ASTIN Bulletin – The Journal of the International Actuarial Association* 23.2 (1993), pp. 213–225. DOI: 10.2143/AST.23.2.2005092 (cit. on pp. 88, 107, 110).

[69] T. Mack. "Which stochastic model is underlying the chain ladder method?" In: *Insurance: Mathematics and Economics* 15.2 (1994), pp. 133–138. DOI: 10.1016/0167-6687(94)90789-7 (cit. on pp. 65, 75, 92).

[70] R. Mahmoudvand and H. Hassani. "Generalized Bonus-Malus Systems with a Frequency and a Severity Component on an Individual Basis in Automobile Insurance". In: *ASTIN Bulletin* 39.1 (May 2009), pp. 307–315. DOI: 10.2143/AST.39.1.2038066 (cit. on pp. 9, 129).

[71] L. Márkus and J. Gyarmati-Szabó. "A Hierarchical Bayesian model to predict belatedly reported claims in insurances". In: *Recent Advances In Stochastic Modeling And Data Analysis*. World Scientific, 2007. Chap. 17, pp. 137–144. DOI: 10.1142/9789812709691_0017. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789812709691_0017 (cit. on p. 67).

[72] L. Martinek. "Analysis of Stochastic Reserving Models By Means of NAIC Claims Data". http://web.cs.elte.hu/~martinek/papers/paper_martinek_stochres_naic.pdf. Manuscript submitted for publication. 2019 (cit. on pp. 4, 68, 103, 104, 126).

[73] L. Martinek and N. M. Arató. “An approach to merit rating by means of autoregressive sequences”. In: *Insurance: Mathematics and Economics* 85 (2019), pp. 205–217. DOI: 10.1016/j.insmatheco.2019.01.008 (cit. on pp. 4, 9, 32).

[74] M. D. Martínez-Miranda, J. P. Nielsen, and R. Verrall. “Double Chain Ladder and Bornhuetter-Ferguson”. In: *North American Actuarial Journal* 17.2 (2013), pp. 101–113. DOI: 10.1080/10920277.2013.793158. eprint: https://doi.org/10.1080/10920277.2013.793158 (cit. on p. 66).

[75] T. Mayuzumi. “A Study of the Bonus-Malus system”. In: *ASTIN Colloquium International Actuarial Association* (1999), pp. 79–94 (cit. on p. 11).

[76] M. Merz and M. V. Wüthrich. “Paid-incurred chain claims reserving method”. In: *Insurance: Mathematics and Economics* 46.3 (2010), pp. 568–579. DOI: 10.1016/j.insmatheco.2010.02.004 (cit. on p. 67).

[77] G. Meyers. “Stochastic Loss Reserving Using Bayesian MCMC Models”. In: *CAS Monograph Series* 1 (2015) (cit. on pp. 3, 66, 103, 104, 107, 108).

[78] G. Meyers. *Thinking Outside the Triangle, Paper presented to the 37th ASTIN Colloquium, Florida*. Paper presented to the 37th ASTIN Colloquium, Florida, 2007 (cit. on p. 102).

[79] G. G. Meyers and P. Shi. “The retrospective testing of stochastic loss reserve models”. In: *Casualty Actuarial Society E-Forum, Summer*. 2011 (cit. on pp. 103, 104).

[80] R. Norberg. “A credibility theory for automobile bonus systems”. In: *Scandinavian Actuarial Journal* 1976.2 (1976), pp. 92–107. DOI: 10.1080/03461238.1976.10405605 (cit. on p. 31).

[81] S. C. Park, J. Lemaire, and C. T. Chua. “Is the Design of Bonus-Malus Systems Influenced by Insurance Maturity or National Culture? — Evidence from Asia”. In: *The Geneva Papers on Risk and Insurance - Issues and Practice* 35.1 (2010), S7–S27. DOI: 10.1057/gpp.2010.37 (cit. on p. 11).

[82] E. S. Pearson. “The probability integral transformation for testing goodness of fit and combining independent tests of significance”. In: *Biometrika* 30.1/2 (1938), pp. 134–148. DOI: 10.2307/2332229 (cit. on p. 82).

[83] K. Pearson. “On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random”. In: *Biometrika* 25.3–4 (1933), pp. 379–410. DOI: 10.1093/biomet/25.3-4.379 (cit. on p. 82).

[84] M. Pigeon, K. Antonio, and M. Denuit. “Individual Loss Reserving with the Multivariate Skew Normal Framework”. In: *ASTIN Bulletin* 43.3 (2013), pp. 399–428. DOI: 10.1017/asb.2013.20 (cit. on p. 102).

[85]P. J. R. Pinheiro, J. M. Andrade e Silva, and M. De Lourdes Centeno. “Bootstrap Methodology in Claim Reserving”. English. In: *The Journal of Risk and Insurance* 70.4 (2003), pp. 701–714. DOI: `10.1046/j.0022-4367.2003.00071.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.0022-4367.2003.00071.x` (cit. on pp. 67, 78).

[86]S. Pitrebois, M. Denuit, and J-F. Walhin. “Fitting the belgian bonus-malus system”. In: *Belgian Actuarial Bulletin* 3.1 (2003), pp. 58–62 (cit. on pp. 18, 32, 53).

[87]G. Quarg and T. Mack. “Munich chain ladder”. In: *Blätter der DGVFM* 26.4 (2004), pp. 597–630. DOI: `10.1007/BF02808969` (cit. on p. 113).

[88]R.A. Rigby and D.M. Stasinopoulos. “A flexible regression approach using GAMLSS in R”. In: *London Metropolitan University, London* (2009) (cit. on p. 47).

[89]M. Rosenblatt. “Prediction and Non-Gaussian Autoregressive Stationary Sequences”. In: *The Annals of Applied Probability* 5.1 (1995), pp. 239–247. DOI: `doi:10.1214/aoap/1177004838` (cit. on p. 35).

[90]M. Rosenblatt. “Remarks on a Multivariate Transformation”. In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472. DOI: `10.1214/aoms/1177729394` (cit. on p. 82).

[91]M. S. Roulston and L. A. Smith. “Evaluating Probabilistic Forecasts Using Information Theory”. In: *Monthly Weather Review* 130.6 (2002), pp. 1653–1660. DOI: 10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2 (cit. on p. 21).

[92]L. J. Savage. “Elicitation of Personal Probabilities and Expectations”. In: *Journal of the American Statistical Association* 66.336 (1971), pp. 783–801. DOI: `10.1080/01621459.1971.10482346`. eprint: `https://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1971.10482346` (cit. on p. 23).

[93]M. J. Schervish. “A General Method for Comparing Probability Assessors”. In: *The Annals of Statistics* 17.4 (1989), pp. 1856–1879. DOI: `doi:10.1214/aos/1176347398` (cit. on p. 23).

[94]M. R. Shapland. “Using the ODP Bootstrap Model: A Practitioner’s Guide”. In: *Casualty Actuarial Society Monograph* 4 (2016) (cit. on p. 78).

[95]P. Shi. “A Multivariate Analysis of Intercompany Loss Triangles”. In: *Journal of Risk and Insurance* 84.2 (2015), pp. 717–737. DOI: `10.1111/jori.12102`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/jori.12102` (cit. on p. 106).

[96]P. Shi and E. W. Frees. “Dependent Loss Reserving using Copulas”. In: *ASTIN Bulletin* 41.2 (2011), pp. 449–486. DOI: `10.2143/AST.41.2.2136985` (cit. on pp. 66, 106).

[97]P. Shi and B. M. Hartman. “Credibility in Loss Reserving”. In: *North American Actuarial Journal* 20.2 (2016), pp. 114–132. DOI: `10.1080/10920277.2015.1109456`. eprint: `https://doi.org/10.1080/10920277.2015.1109456` (cit. on p. 109).

[98] P. Shi, S. Basu, and G. G. Meyers. “A Bayesian Log-Normal Model for Multivariate Loss Reserving”. In: *North American Actuarial Journal* 16.1 (2012), pp. 29–51. DOI: 10.1080/10920277.2012.10590631. eprint: https://doi.org/10.1080/10920277.2012.10590631 (cit. on p. 66).

[99] F. A. Sloan. “Experience Rating: Does It Make Sense for Medical Malpractice Insurance?” In: *The American Economic Review* 80.2 (1990), pp. 128–133 (cit. on p. 8).

[100] G. J. Székely and M. L. Rizzo. “A new test for multivariate normality”. In: *Journal of Multivariate Analysis* 93.1 (2005), pp. 58–80. DOI: 10.1016/j.jmva.2003.12.002 (cit. on p. 85).

[101] C. I. Tan, J. Li, J. S-H. Li, and U. Balasooriya. “Optimal relativities and transition rules of a bonus-malus system”. In: *Insurance: Mathematics and Economics* 61.C (2015), pp. 255–263. DOI: 10.1016/j.insmatheco.2015.02.001 (cit. on p. 31).

[102] G. C. Taylor. *Loss Reserving: An Actuarial Perspective*. Vol. 21. Huebner International Series on Risk, Insurance and Economic Security Series. Springer Science & Business Media, 2000. DOI: 10.1007/978-1-4615-4583-5 (cit. on p. 65).

[103] G. C. Taylor and F. R. Ashe. “Second moments of estimates of outstanding claims”. In: *Journal of Econometrics* 23.1 (1983), pp. 37–61. DOI: 10.1016/0304-4076(83)90074-X (cit. on p. 75).

[104] L. Tee, M. Käärik, and R. Viin. “On Comparison of Stochastic Reserving Methods with Bootstrapping”. In: *Risks* 5.1 (2017), p. 2. DOI: 10.3390/risks5010002 (cit. on pp. 66, 84).

[105] L. Tremblay. “Using the Poisson Inverse Gaussian in Bonus-Malus Systems”. In: *ASTIN Bulletin* 22.1 (1992), pp. 97–106. DOI: 10.2143/AST.22.1.2005129 (cit. on p. 10).

[106] G. Tzougas, S. Vrontos, and N. Frangos. “Optimal Bonus-malus Systems Using Finite Mixture Models”. In: *ASTIN Bulletin* 44.2 (2014), pp. 417–444. DOI: 10.1017/asb.2013.31 (cit. on pp. 8, 31).

[107] R. J. Verrall. “A Method for Modelling Varying Run-off Evolutions In Claims Reserving”. In: *ASTIN Bulletin – The Journal of the International Actuarial Association* 24.2 (1994), pp. 325–332. DOI: 10.2143/AST.24.2.2005074 (cit. on p. 65).

[108] J. F. Walhin and J. Paris. “Using Mixed Poisson Processes in Connection with Bonus-Malus Systems”. In: *ASTIN Bulletin* 29.1 (1999), pp. 81–100. DOI: 10.2143/AST.29.1.504607 (cit. on p. 10).

[109] M. V. Wüthrich. “Accounting Year Effects Modeling in the Stochastic Chain Ladder Reserving Method”. In: *North American Actuarial Journal* 14.2 (2010), pp. 235–255. DOI: 10.1080/10920277.2010.10597587. eprint: https://doi.org/10.1080/10920277.2010.10597587 (cit. on p. 66).

[110] M. V. Wüthrich. “Claims Reserving Using Tweedie's Compound Poisson Model”. In: *ASTIN Bulletin* 33.2 (2003), pp. 331–346. DOI: 10.1017/S0515036100013490 (cit. on p. 106).

[111]M. V. Wüthrich and M. Merz. *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, 2008. DOI: 10.1002/9781119206262 (cit. on pp. 2, 65, 66, 70–72, 74, 78).

[112]K. C. H. Yip and K. K. W. Yau. “On modeling claim frequency data in general insurance with extra zeros”. In: *Insurance: Mathematics and Economics* 36.2 (2005), pp. 153–163. DOI: 10.1016/j.insmatheco.2004.11.002 (cit. on pp. 9, 33, 62).

[113]B. Zehnwirth. “Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital”. In: *Variability in Reserves Prize Program Papers*. Vol. 2. Casualty Actuarial Science Forum. 1994, pp. 447–606 (cit. on p. 75).

List of Figures

List of Tables

Summary

We have researched mathematical applications in non-life insurance with particular focus on experience rating, bonus–malus systems and stochastic claims reserving.

The market of third-party liability insurance is large and insurance institutions try to optimise their expected claim estimation methods. We have shown on different models how certain information and time-dependence affect such approximations. The ranking framework has been constructed in an algorithmic manner with scores. As a subsequent step, we have scrutinised claim processes with a more general premise, assuming that the underlying process is a stationary Markov chain. As a result, we have defined a new merit rating scheme which is structurally different from the schemes currently in application. This autoregressive process-based premium adjustment has been put into context by comparing it with existing models. This comparison has been performed by calculating key metrics such as elasticity, coefficient of variation, financial equilibrium, etc., also with analytical formulas in respect of the new model.

Non-life insurance undertakings put a substantial effort into understanding the nature of their outstanding liabilities. Multiple versions of stochastic reserving models emerged in the past two decades as a result. We have proposed an algorithmic manner of model selection from competing models and looked at reserving methods as probabilistic forecasts. This simultion-based ranking framework involves metrics which are partly new in the actuarial reserving context: probability scores, probability integral transform, empirical coverage and sharpness. Insurance firms do not or rarely publish their claim histories, hence, drawing pseudo-realities have been necessary to test the theoretical models. These metrics have shown from several angles how one model outperforms the other. Next to the purely simulation results, reserving models have also been tested on real-life data provided by hundreds of companies, and on a policy-level data set from one company. We have proposed two new stochastic methods, the credibility bootstrap and the collective semi-stochastic, which exploit observations from a collective of institutions.

We have used `R` to generate a vast amount of simulations, tables and graphs. It also helped us in the preliminary phase of the research to define the relevant conjectures.

Összefoglalás

Kutatásunk a nem-életbiztosítás matematikai alkalmazásaira összpontosít: a tapasztalati díjszámításra, bónusz–málusz rendszerekre és sztochasztikus kártartalékolásra.

Mivel a kötelező felelősségbiztosítások piaca jelentős méretű, a biztosítók nagyban törekszenek a várható kárbecslési módszerek tökéletesítésére. Néhány modellen megmutattuk, milyen hatással vannak a rendelkezésre álló információk és az eltelt idő a becslésekre. Szkórok használatával algoritmikus úton készítettünk becslési módszerek rangsorolására alkalmas keretrendszert. Ezt követően a kárfolyamatokat egy általánosabb megközelítésben vizsgáltuk azzal a feltétellel, hogy a mögöttes kárfolyamat stacionárius Markov-lánc. Ennek eredményeként egy új tapasztalati díjkalkulációs rendszert dolgoztunk ki, amely szerkezetében eltér a jelenleg alkalmazottaktól és amelyben a díjak kiigazításáért egy autoregresszív folyamat felel. Ezt létező bónusz–málusz rendszerekkel való összehasonlítások révén helyeztük kontextusba. Számításba vettünk néhány fontos metrikát, az elaszticitást, szórási együtthatót, pénzügyi stabilitást stb., az új modellben esetenként analitikus formulákat is használva.

Kintlévő kötelezettségeik természetének megértésére a nem-életbiztosítási intézmények jelentős hangsúlyt fektetnek. Számos sztochasztikus kártartalékolási modell született az elmúlt két évtizedben. Egymással versengő modellek közül algoritmikus úton javasoljuk kiválasztani a legjobbat, miközben valószínűségi előrejelzésként tekintünk ezen tartalékolási módszerekre. A bevezetett szimulációkon alapuló rangsorolás részben olyan metrikákat is felhasznál, amelyek eddig idegenek voltak az aktuáriusi tartalékolás számára. Ezek a valószínűségi szkórok, a valószínűségi integrál-transzformáció, az empirikus lefedettség és az élesség. Mivel a biztosítók csak elvétve publikálják tételes kártörténeteiket, szimulált kártörténetek útján ellenőriztük az elméleti modelleket. A tartalékolási modelleket az említett mérőszámok mérik össze teljesítményük számos szempontjából. A szimulációk mellett valódi káradatokon is végeztünk összehasonlítást, egyrészt egy több száz vállalattól gyűjtött kárkifutási adatbázis, másrészt egy egyetlen cégtől származó szerződésszintű kártörténet alapján. Az eredmények értelmezése fontos és részletes értekezést igényel. Két új sztochasztikus tartalékolási módszert javasoltunk, a *credibility bootstrap* és a kollektív szemi-sztochasztikus módszereket, amelyek több intézménytől gyűjtött megfigyeléseket vesznek egy időben figyelembe.

Az `R` programozási nyelv segítségével generáltuk a kutatás során használt nagy mennyiségű szimulációt, táblázatokat, ábrákat, illetve a kutatás kezdeti fázisában nagy segítségünkre volt a lényegesebb sejtések megfogalmazásában.

Index

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

ADATLAP a doktori értekezés nyilvánosságra hozatalához

I. A doktori értekezés adatai

A szerző neve: Martinek László

A doktori értekezés címe és alcíme:
Experience Rating and Stochastic Reserving in General Insurance
(magyarul: Tapasztalati díjszámítás és sztochasztikus tartalékolás a nem-életbiztosításban)

A doktori iskola neve: Matematika Doktori Iskola

A doktori iskolán belüli doktori program neve: Alkalmazott Matematika Doktori Program

A témavezető neve és tudományos fokozata: Arató Miklós, egyetemi docens, CSc

A témavezető munkahelye: ELTE TTK, Valószínűségelméleti és Statisztika Tanszék

MTA Adatbázis-azonosító: 10059884

DOI-azonosító[1]: 10.15476/ELTE.2019.069

II. Nyilatkozatok

1. A doktori értekezés szerzőjeként[2]

a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a ……Természettudományi Kar, Dékáni Hivatal, Doktori, Habilitációs és Nemzetközi Ügyek Csoportjának ügyintézőjét…..., hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.

b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[3]

c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (*dátum*)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[4]

1 A kari hivatal ügyintézője tölti ki.
2 A megfelelő szöveg aláhúzandó.
3 A doktori értekezés benyújtásával egyidejűleg be kell adni a tudományági doktori tanácshoz a szabadalmi, illetőleg oltalmi bejelentést tanúsító okiratot és a nyilvánosságra hozatal elhalasztása iránti kérelmet.
4 A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a minősített adatra vonatkozó közokiratot.

d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.[5]

2. A doktori értekezés szerzőjeként kijelentem, hogy

a) a ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;

b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.

3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: Budapest, 2019. március 28.

Martinek László
a doktori értekezés szerzőjének aláírása

5 A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a mű kiadásáról szóló kiadói szerződést.